Psychological Bulletin

CONTENTS

Published Bimonthly by the
American Psychological Association

VOL. 53, No. 2 MARCH, 1956

The Psychological Bulletin contains evaluative reviews of research literature and articles on research methodology in psychology. This JOURNAL *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Manuscripts should be sent to this address even during the editor's absence on leave. Manuscripts will be acknowledged at this address and the carbon copies forwarded to the editor. However, beginning on August 1, 1955 and until August 1, 1956 *correspondence* should be addressed to the editor at the Department of Psychology, American University, Beirut, Lebanon.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 [No. 4, Part 2], 389–449). Special attention should be given to the section on the preparation of the references (pp. 432–440), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate*. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Reprints. Fifty free offprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis offprints.

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage under the provisions of Sec. 34-40 Par. (D) provided for in Section 538, act of February 28, 1925, authorized August 6, 1947. Printed in U.S.A.

Vol. 53, No. 2 March, 1956

Psychological Bulletin

THEORIES OF VISUAL ACUITY AND THEIR PHYSIOLOGICAL BASES[1]

JOHN L. FALK
University of Illinois

The Problem of Visual Acuity

Visual acuity is defined as the reciprocal of the minimum visible angle measured in minutes of arc. The assumption is that absolute size or distance of the test-object is not important, but only the angle it subtends at the eye. There are some indications in the literature that acuity may actually be a function of distance even when the visual angle remains constant, when the distances are less than one meter or so (**13**, **72**). However, this is explicable in terms of the varying efficiency of accommodation.

The most common acuity measures are those dealing with the ***minimum separable*** and the ***minimum visible.*** Minimum separable tests deal with the perception of a small gap between two parallel bars. Minimum visible tests present a single, fine line for discrimination upon a homogeneous background. Actually, no clear distinction can be drawn between these two categories, for as we enlarge the parallel bars of the minimum separable test, we are at some point no longer dealing with this kind of test but with a minimum visible test; the bars have become homogeneous background and the gap the sole figure (the line to be discriminated). Coincident with the above transformation will be an improvement in acuity or the ability to resolve the stimulus figure. There are, in fact, a variety of ways for measuring visual acuity, more often than not yielding results which are incommensurable with one another. For example, acuity is maximal for a single line, it is somewhat less for stereoscopic and vernier methods of measurement, while the detection of the minimum separable for two bright bars is distinctly less well performed.

The reader should be clear as to the difference between acuity and sensitivity. *Sensitivity* refers to the capacity of the organism to respond to small values of photic intensity, while by *acuity* is meant the capacity to distinguish (resolve) very fine or very close details. In general, the sufficient stimulus for the rod receptors in terms of radiant energy is lower than that associated with the cones, so that except for the red end of the spectrum, where sensitivity is about equal, rods are more sensitive than cones. Chapanis (**13**) refers to experiments which indicate that while a given level of dark adaptation allows light of a certain luminance to be perceived (sensitivity) it does not

[1] The author wishes to thank Dr. Peter Milner, McGill University, for helpful comments.

insure the ability to discriminate forms at this luminance level (acuity). Also, under scotopic conditions the retinal region of maximum sensitivity does not correspond to the region of maximum acuity. Later we shall consider certain structural features which contribute to acuity and sensitivity.

Many factors have been shown to influence the degree of resolution: the luminance and size of the background, surround luminance, wavelength, pupil-size, exposure time, retinal area stimulated, state of adaptation, etc. One is first inclined to attempt an explanation of acuity in terms of the projection of a geometrical image upon the retina. This interpretation must be rejected for two cardinal reasons. First, owing to its distortion the retinal image is far from being a pictorial replica of the visual world; and second, the retinal receptor mosaic, fine as it is in the central fovea, is not fine enough to account for the degree of visual resolution possible. However, as we shall presently indicate, a finer receptor "grain" might not give rise to greater acuity at all.

Let us discuss these factors of image blurredness and receptor size in somewhat greater detail. The image is brought to focus on the light-sensitive retina by the cornea and the delicately adjusted accommodation of the lens. The pupil responds to the luminance of the field of view by widening or narrowing. At this point any analogy to a camera must be abandoned. There is a spread of light from the geometrical image of a bright object to its dark background (or into a dark object from a brighter background) owing to the diffraction of light by the pupillary aperture, as well as to the spherical and chromatic aberrations. For a pupil size of 2.5 mm. or less, the spread of light is attributable almost entirely to diffraction. For larger pupil sizes, the aberrations become important factors. Acuity has been found to increase up to a pupillary diameter of 2 mm., remain constant from 2 to 4 mm., and then decrease owing to chromatic aberration.

There are still further respects in which the eye is quite unlike a camera. It is not light-tight, and allows light to pass through the sclerotic coat surrounding the cornea. There is a further diffraction and scattering of light in the internal media of the eye. Light must pass through retinal blood vessels, nerve fibers, and cell bodies in order to reach the rods and cones. And there is reflection from the formed image on the retina to all other parts of the retina. Bartley (**6**, p. 58) claims that under many conditions "the level of stray illumination is a considerable fraction of the intensity of the image itself." Further, "with a single disk [as image], the major factor producing the electroretinogram is fluctuation in stray light. When stray light is constant, the conditions for producing a series of waves in the electroretinogram are absent" (**7**, p. 926).

The photographic plate has an even distribution of light-sensitive substance upon its surface, while the density of rods and cones varies according to retinal region. There are not only regional variations in the type, size, and spacing of the retinal receptors, but also the initial retinal "grain" is substantially altered by various synaptic relationships (e.g., the funneling of a group of rods onto one ganglion cell). Again, the image, the points of which are composed of small "blur circles," is focused exactly only for a small central region, the more peripheral areas of the image being somewhat out of focus (**17**). For cone vision it has been

found that light passing into the eye obliquely at the edge of the pupil is a less effective stimulus than light entering through the center of the pupil. This directional sensitivity is termed the Stiles-Crawford effect and has been explained (**57, 58**) by considerations of cone shape, refractive index, and locus of photosensitive pigments. The adaptive range of the eye is over 1,000,000 to 1, but pupil area varies only through a 16-to-1 range. Clearly, adaptation must be explained by some additional factors. Also, the Stiles-Crawford effect has shown that even this pupil range has less control over alterations in retinal illuminance than was formerly assumed (**54**, p. 403).

To this complex picture of illuminance distribution we may add the fact that the eyes are in a state of constant fluttering motion termed *physiological nystagmus*. This fine, rapid tremor is to be distinguished from the larger saccadic "flicks" and slow "drifts" (**18**). Physiological nystagmus occurs since the eyeball is held in balance between pairs of antagonistic muscles, and unlike the larger movements, it is not binocularly coordinated. The measurements of tremor amplitude do not agree too well among the various experimenters, but the effect itself is well established.

Thus, despite accommodation of the lens for maximum acuity, we find that the image delivered to the retina is blurred and distorted by many factors: diffraction at the pupil, spherical and chromatic aberration, passage of extraneous light through the sclerotic coat, diffraction and scattering of light in the internal media, the imposition of blood vessels and neural tissue between incoming light and the receptors, reflection from the formed image, differential densities of receptor elements, lack of sharp focus except at the image center, the Stiles-Crawford effect, and physiological nystagmus.[2]

We must now consider the second factor arguing against the interpretation of acuity by the geometry of images. This is the set of structural characteristics of the retinal mosaic in the light of acuity performances. According to the geometrical interpretation, for resolution of the details of a pattern to occur, the image formed on the retina must not have a fineness of grain exceeding that exhibited by the diameter of the individual retinal receptors. But Hecht and Mintz (**41**) found that a thin, dark line subtending 0.5″ of visual arc could be resolved. This is roughly equivalent to seeing a 1/16-inch wire at a distance of half a mile. By the geometry of the situation, the image of this line (uncorrected for diffraction, etc.) would cover only about 1/40 of the diameter of the smallest foveal cones. The smallest cones in the fovea centralis are 1 to 1.5μ in diameter and therefore subtend a visual arc of about 12″ to 18″ which is considerably larger than the 0.5″ minimum visible found by Hecht and Mintz. How is it possible for the eye to discriminate details finer than its own mosaic? One reason is that the 0.5″ geometrical image is a fiction; the sharpest edge will spread its blurred image over at least four or five foveal receptors. Also, it has been suggested that within certain

[2] Some additional factors producing distortion of the retinal image are given by Tschermak-Seysenegg (**76**, pp. 7–10). Lest the amount and complexities of retinal image distortion appear too great for fine resolution ever to occur, it should be noted that the eye possesses certain mechanisms which help in correcting the aberrations: the differential curvature and density of the lens compensates for spherical aberration, while its yellow color reduces chromatic aberration by eliminating the ultraviolet end of the spectrum.

limits, rate of firing may be "transposed . . . into an 'angular value,' that is, interpreted dimensionally: low intensities as small, high intensities as large, angular sizes" (**64**, p. 429).

Summarizing, we may say that there are at least two problem paradoxes in the explanation of visual resolution. The image formed upon the retinal mosaic appears to be a rapidly vibrated blur. The mosaic itself, tight-packed with fine cells in the foveal region, is still too coarse geometrically to account for the fineness of visual acuity. However, in the conjunction of these two paradoxes investigators have seen various ways of solving both of them. These lines of thought will be discussed under the somewhat arbitrary headings of "static theories" and "dynamic theories."

There are certain major factors in the problem of visual acuity which should be kept in mind, since it is interesting to see how the various theorists handle them: (*a*) the blurred character of the retinal image, (*b*) the continuous motion of the image due to tremor, (*c*) the diameter and spacing of the visual cells, (*d*) the modes of connection of the visual cells to the optic nerve, and the way in which the rest of the visual system is assumed to affect acuity.

The Size and Distribution of Retinal Receptors

This section is intended to give the reader the barest outline of anatomical features in the retina required to comprehend the theories to be discussed. Other material of this nature will be discussed in the context of the theory utilizing it.

The structural limitations of visual resolution, aside from the blurring of the image, begin with the size and separation of the receptor elements in the retina. Since rods are absent from the fovea centralis and are not implicated in photopic acuity they will be ignored in preference to the cones in this discussion.

> The thinnest cones are present in the very center of the fovea, where their thickness, in Man, is reduced to almost 1μ, corresponding to a visual angle of approximately 12″ to 15″ of arc. . . . The central territory, where the cones are almost uniformly thick, measures approximately 100μ across, corresponding to 20′ . . . of arc. . . . It contains approximately 50 cones in a line. This area seems to be not exactly circular but elliptical, with the long axis horizontal, and may contain altogether 2000 cones . . . here practically every cone is individually linked with the ganglion cells by means of its own "private" midget bipolar . . . it represents an independent functional unit . . . the size of each of the 2000 receptor-conductor units measures, on the average, 24″ of arc (**64**, pp. 425–26).

At the edge of the fovea, the number of cones to 100μ (20′ of arc) is reduced to 30, and their size increases to 40″ of arc. This trend continues as we move further and further into the periphery, the cones becoming larger and more sparse.

The cones of the fovea are mostly hexagonal, and " . . . in certain directions there are rows arranged in straight or nearly straight lines; in other directions this arrangement is less pronounced or is altogether absent. But even where regular, the rows are straight on short stretches only, soon becoming a tangle of rows and groups of cones oriented differently" (**64**, p. 427).

The cones are separated by thin partitions of neuroglia, the thickness of which is less than the diameter of the thinnest cone. These partitions vary from $\frac{1}{3}\mu$ or $\frac{1}{5}\mu$ in the very center to $\frac{1}{2}\mu$ (6″ of visual arc) in the periphery.

The fovea itself consists of a flat, shallow pit in the retinal surface in primates. The particular visual purpose served by this arrangement is a

subject of debate among experts. For two divergent views see Polyak (**64**, pp. 209–210) and Walls (**77**, pp. 183–184).

As might be suspected, the visual acuity of particular retinal areas is closely associated with the cone density there (see **64**, Fig. 99). Recent evidence (**44**, **57**) indicates that our highest acuity is confined to a central region 35μ in diameter (7′ of arc). Just beyond this tiny region (26 cones across, containing a total of approximately 600 cones) there is a 5 per cent decrement is acuity.

Static Theories

Intensity Discrimination Theory (Hartridge)

Hartridge (**35**) presented details of his theory of visual acuity in 1922, and apparently still holds to the main outlines of it, although admitting that it was wrong in its quantitative aspects (**36**). The point of departure consists in considering the acuity task to be a particular case of intensity discrimination. The blurred retinal image is regarded as a distribution of intensities. With a minimum visible bright line as test-object, the cone at the center of the image will receive light of greater intensity than those cones on either side of it. This threshold intensity difference was calculated by means of Rayleigh's equations, which take account of the diffraction and aberration characteristics of light, and was found to approximate 10 per cent. Thus, the eye actually is considered to turn the blurredness of the retinal image to its advantage. In contrast to the geometrical interpretation of visual acuity, Hartridge's theory posits that "... the diffraction pattern of an object formed on the retina must not exhibit differences of intensity which are below the threshold of difference perception of the retinal receptors" (**36**). As shown in Figure 1, the only requirement for resolution of the minimum separable gap is that the illuminance incident at the central cone row be perceptibly different from that received by rows on either side.

Intensity Discrimination Theory (Hecht)

Hecht starts with the fact that minimum separable acuity varies with the intensity of illumination, and utilizing the analogy of the photographic plate, states that the fineness of detail which the retina can register is dependent upon the density of active receiving elements. By combining this conception of the visual system with the assumption that the sensitivity of receptors is randomly distributed,[3] Hecht (**39**) is able to give an account of the increase in acuity with increased field luminance. The integral of the Gaussian distribution curve is an S-shaped function which is fitted to König's intensity-acuity data. According to Hecht (**38**, **39**) an increase in photic intensity will give rise to an increased number of retinal elements which are functional, thereby effecting a reduction in the average center-to-center distance of active elements and increasing the resolving power of the surface. Visual acuity, then, "varies directly with the number of functional rods or cones in a unit area of illuminated retina" (**38**).

Like Hartridge, Hecht considers the image on the retina as a pattern of intensities. If one row of cones is subjected to an intensity difference amounting to only 1 per cent with respect to adjacent rows, then this difference will be perceived. Since

[3] This is assumed for both rod and cone populations, although the rods are regarded as having lower thresholds in general.

only one cone row is differentially affected, the line (in the case of the minimum visible) will be sharp rather than fuzzy[4] (**41**). Also the intensity difference between the line and the background retinal illuminance is imposed on different "center rows" of cones in a short time interval owing to eye movements.

Finally, Hecht and Mintz (**41**) note that an intensity difference of 1 per cent is near the limit of differential intensity discrimination. They suggest that the just resolvable visual angle varies with light intensity in the same fashion as the capacity to discriminate intensity differences varies with intensity.

Critique of Hecht's theory. In view of the wide influence which this theory has exerted upon current conceptions of visual acuity, it may be well to indicate some of the evidence which fails to support it.[5]

Byram (**12**) has pointed out that the equations which Rayleigh derived for the distribution of light intensity on the retina in the case of a long, black line on a uniformly bright field are valid only for a rectangular aperture. Since the pupil is circular, the calculations of Hecht and Mintz err by 15 per cent. Hartridge's error amounts to some 200 per cent.

Lythgoe (**49**) has shown that the classical, sigmoid intensity-acuity curve obtained by König errs at both the high and low ends owing to a failure to maintain a surround luminance comparable to that of the test-object background. Lythgoe was able to reproduce König's results at the high end (a leveling off or slight falling off of acuity at high task

[4] From what we have already noted above concerning Polyak's denial of the existence of straight "cone rows" it would seem necessary to explain the sharpness of fine resolution upon some other basis.

[5] The reader who desires additional critical material on Hecht's theory should consult the excellent review articles by Senders (**70**) and Walls (**78**).

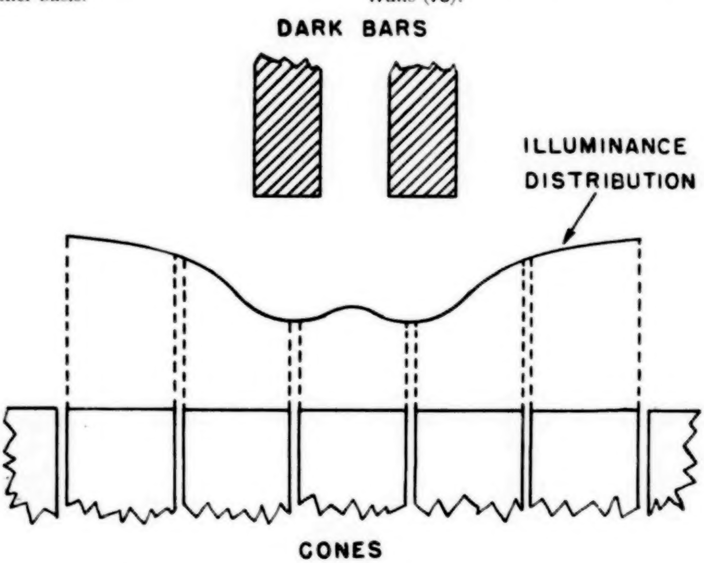

FIG. 1. AN ILLUSTRATION OF THE RETINAL BASIS OF MINIMUM SEPARABLE DISCRIMINATION ACCORDING TO HARTRIDGE'S THEORY (MODIFIED FROM WALLS [**78**])

luminances) with inadequate surround luminance; but when adequate surround luminance was maintained, visual acuity continued to increase with increasing luminance of the acuity task. Thus, any theory which views visual acuity as explicable solely in terms of brightness discrimination (e.g. Hecht's) is faced with the following problem: the differential brightness threshold ceases to improve at high brightnesses (and perhaps deteriorates) while acuity continues to increase up to the highest luminance value used by any experimenter! Moon (**54**) describes a study by Eguchi in which acuity continued to rise up to values of luminosity twice as great as that of white paper in direct noon sunlight. Moon comments further that ". . . neither Lythgoe nor Eguchi succeeded, at the highest values of [luminosity], in illuminating the surround to the luminosity of the test-object background. Thus we have no proof that the straight-line relation may not continue upward considerably beyond the point at which even the Eguchi tests indicate a bend" (**54**, p. 435).

Hecht's theory applies to König's curve which is flat-topped and does not correctly represent the intensity-acuity relation under more adequately controlled conditions. Likewise, at the low-intensity end of the curve Lythgoe (**49**) obtained no flattening-out, a result which indicates that the sharp discontinuity of the rod and cone limbs of the curve may not actually occur.

The assumption of a random distribution of thresholds for retinal elements has been called into question by many investigators. Few have been willing to accept the view that cone thresholds vary so widely. In the light of the results of Lythgoe and Eguchi one would be forced to conceive of the retina as a sensitive surface in which the thresholds of some of its elements have never been reached. Fine as the retinal grain is, the theorist is hard put to explain the greater fineness of acuity performances, let alone ruling out a considerable part of the mosaic for having thresholds so high that they would seldom (if ever) be attained.

Hecht has been criticized for all but disregarding an important dimension of neural action, namely, that the *frequency* of neural discharge varies with stimulus intensity. Although there is undoubtedly some variation among cone thresholds, the most likely candidate for the representation of intensity level is the relative discharge rate passed on to optic nerve fibers rather than changes in the absolute number of receptors excited (**3**, **33**). The receptor responds in a *graded* rather than in the all-or-none fashion Hecht assumes.

Actually, the assumption of widely separated thresholds in percipient-element populations gives rise to some rather peculiar implications. For example, consider the discrimination of a small, dark spot on a moderately illuminated background. The dark area will produce an effect of illuminance decrement on a number of cones and the dark spot will be perceived. But according to Hecht's theory there must be many such areas under moderate illumination conditions since the thresholds of many of the cones have not yet been reached. The homogeneous background would then appear as a rather spotty, blotchy expanse in which the retinal representation of the true dark spot would be indistinguishable from small clusters of elements with higher thresholds. Needless to say, such circumstances do not exist.

The Hecht theory is unifactorial—only intensity is assumed to affect

acuity. Other factors influencing acuity performances have been neglected (e.g., size and shape of the test-object). The fact that increasing luminance gives rise to only a slight change in vernier acuity (aligning power) presents a further difficulty for the theory (**1**, **9**). Both Berger (**8**) and Ogle (**61**) note that with two point sources on a dark background the minimum angle of resolution increases (resolving power decreases) as the luminance of the points is increased. Ogle (**61**) has explored this effect using various background luminances and concludes: " . . . it would appear that Hecht's theory for the increase of visual acuity with illumination does not apply, because the MAR [minimum angle of resolution] was apparently a function of luminance ratio or contrast *only*."[6]

Finally, O'Brien and O'Brien (**60**) have subjected Hecht's assumption of large, fixed threshold differences among receptor elements to a fairly direct test. Hecht's theory requires a variation among cone thresholds greater than one thousandfold. These investigators, using a double star test-object, contrived to concentrate approximately three-quarters of the luminous flux of the star image upon only three foveal cones. Considering diffraction, etc., this degree of concentration is probably close to the maximum possible for acuity tests. At surround luminance from .01 to 100 foot-lamberts, "the star illumination necessary for visibility of both stars was never more than twice the illumination at which neither was visible. It is concluded that the small variation among foveal cones here observed is inconsistent with the Hecht theory."

Dynamic Theories

The "dynamic" theories of visual acuity differ from the "static" theories in one important respect: while Hartridge and Hecht give scant attention to eye movements, the physiological nystagmus (continuous, fine ocular tremor) is given an important role in augmenting resolution performances by certain theorists.

The Theory of Weymouth et al.

This theory is an attempt to explain how vernier acuity operates with such remarkable sharpness and accuracy. Hair lines appear straight and clear despite the irregular orientation of "cone rows," and vernier acuity is accurate to 2", or somewhat less than one-tenth of the width of a central foveal cone. Weymouth and his co-workers (**4**, **82**) make use of two assumptions which were held by Hering: first, that each cone has a local sign, and second, that physiological nystagmus contributes to acuity. Actually, the second of these is not an assumption in the strict sense for these investigators, since it is derived from some more basic aspects of the situation at the retina. If two adjacent cones are stimulated they are assumed mutually to affect each other's local sign, the net effect being to shift the pooled local signs to some intermediate position. Thus, nystagmus would give rise to a process of successive averaging, the product of which is termed "retinal mean local sign."

The following, then, is a picture of vernier resolution as it would be ac-

[6] Parenthetically, it is worth mentioning that visual acuity is a complex function of a host of conditions. We choose to view movies in the dark where our acuity is poorer only because the contrast conditions are so much better (**48**, p. 260). But on the other hand, Ogle (**61**) has shown that visual acuity *deteriorates* with increased point luminance on a constant background (increased contrast) and improves with increased luminance of background when the point luminances are constant (decreased contrast).

counted for by this theory. Any cone "touched" by light is assumed to be stimulated (**4**). Consider an edge between the light and dark portions of a visual field. The cones which fall along the retinal image of such an edge would form a staggered and irregularly spaced line, as photomicrographs reveal. The mutual influence of the local signs of such cones will serve to straighten this edge somewhat, and the horizontal nystagmus movements sweeping this diffraction edge over a wider strip of cones will add to this dynamic averaging process. Since equal numbers of stimulated cones will come to lie in all parts of the strip swept out by the diffraction edge under the control of nystagmus, the center of this strip will represent the "center of gravity" of all the stimulated cones.

> The average or mean of these points, which determines the local sign of the straight line, is therefore *not restricted to such units as intercornal distance or cone diameter* but may be accurate to a small fraction of these units just as the mean of a number of measurements made in inches may be accurate to a small fraction of an inch (**4**, italics mine).

For the perception of the vernier offset to attain maximum accuracy the lines must be long enough so that a sufficiently large sample of cones may be averaged. It is a fact that increasing the line length up to a certain limit does aid vernier acuity. According to this view, acuity is relatively independent of the size and spacing of the retinal elements, the dynamic averaging process functionally producing a finer "grained" mosaic for line representation and alignment.

The Theory of Marshall and Talbot

The aim of these authors is to develop a theory of visual acuity based upon factors of the nervous system which have presented embarrassing complications to past theories rather than helpful mechanisms. This is the only current acuity theory which involves the anatomy of the visual projection system up to Brodmann's Area 17, as well as many neurophysiological phenomena ignored by previous theories. Just as the blurring of the retinal image by diffraction, aberrations, etc. was utilized by the "static" theories to explain the extreme fineness of acuity, Marshall and Talbot (**53**) carry the process on several steps by taking account of various neurophysiological phenomena such as facilitation, summation, the neural recovery cycle, overlap and multiplication of pathways.

First, let us consider two anatomical factors of great importance to the theoretical presentation: reciprocal overlap and multiplication of pathway. The principle of reciprocal overlap is easily grasped by relating the diagrammatic representation of Figure 2 to the following description: "In the cat, optic tract endings in the geniculate divide into several branches and as many as 40 ring-shaped boutons have been seen on single radiation cells which may come from as many as 10 optic tract fibers" (**53**). Multiplication of pathway simply refers to the fact that in the monkey, for example, "the optic tract fibers seem to divide into 5 or 6 branches, and each branch ends in a bouton which makes contact with a different radiation cell, resulting in multiplication of the transmitting path" (**53**). Such multiplication, as the visual system ascends to the cortex, provides a probability-distribution of cortical cells for each foveal cone. Each receptor, then, may activate different paths to different cortical cells at different times. This expansion or refinement of "grain" from retina to cortex is independent of the functional and anatomical

spread due to reciprocal overlap. While it is admitted that overlap is minimal for neurons ascending from the fovea centralis and is more characteristic of the periphery, it is nevertheless maintained on both anatomical (**64**, p. 430) and experimental (**30**) grounds that possibilities for such interaction exist.

As the retinal receptors are vibrated back and forth across the contours of the diffraction image of a test-object by the small-amplitude, high-frequency action of physiological nystagmus, they are subjected to rapidly changing illuminance values. Not only the *amplitude* of the illuminance change but also its *rate* of change will determine the magnitude of receptor outputs. The experiments of Hartline and Graham (**33**) on the photoreceptors of *Limulus* have shown rate of stimulus onset as well as intensity to be an important determinant of high discharge frequency. Further, this photoreceptor initially discharges at a high frequency when stimulated but rapidly adapts to a lower discharge rate despite the constancy of the stimulus. Ocular tremor, then, would serve to maintain a high neural discharge rate by imposing sudden illuminance changes to many receptor "rows" over frequently occurring cycles. Nystagmus is assumed to fluctuate "with approximately a statistical occurrence of various amplitudes" (**53**). Thus the small illuminance difference presented by a hair-line test-object might be made liminal by the rapidity of its onset over sets of receptors which are stimulated intermittently, thus by-passing the effects of neural accommodation. As Marshall and Talbot (**53**) remark: "This essential discontinuity of stimulation is necessary for continuous vision through a fatiguable system."

We are now in a position to illustrate how the above factors, in combination with the excitability changes of the neural recovery cycle, might work to produce certain gradients of excitation which become progressively refined and sharpened as they are passed on to higher synaptic levels. Confining our discussion to the retinal level for the present, we shall consider three factors whose salient features and modes of operation have been described above: (*a*) the intensity distribution (cf. diffraction, etc.) of the retinal image, termed the *diffraction image*, (*b*) physiological nystagmus, and (*c*) rate of illuminance change.

In Figure 3 the intensity distribu-

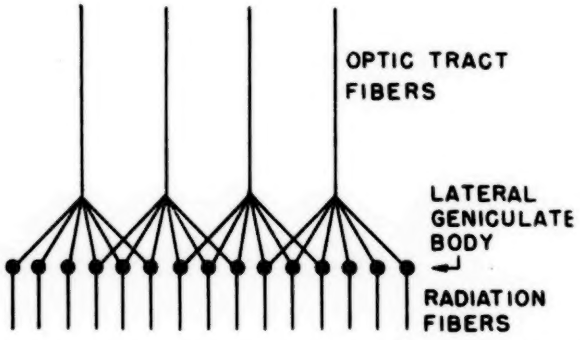

FIG. 2. A DIAGRAMMATIC REPRESENTATION OF RECIPROCAL OVERLAP IN THE LATERAL GENICULATE OF THE CAT

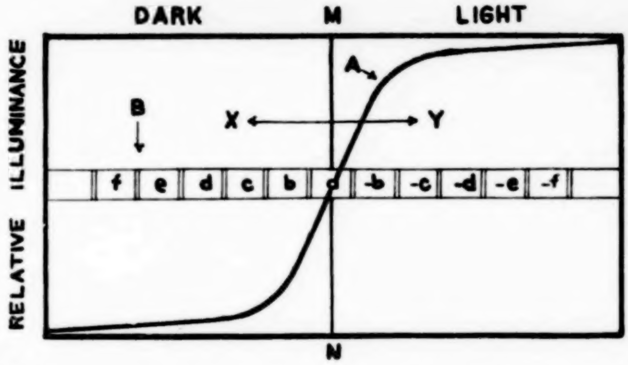

FIG. 3. THE DISTRIBUTION OF ILLUMINANCE ON THE RETINA ACROSS THE GEOMETRICAL BOUNDARY (MN) SEPARATING LIGHT AND DARK HALVES OF A FIELD (MODIFIED FROM JONES AND HIGGINS)

tion of the diffraction image of a border between the light and dark sides of a bipartite field is shown (curve A). The geometrical border is MN. This figure is based upon one given by Jones and Higgins (**44**) who apply some of the conceptions of Marshall and Talbot in their treatment of photographic granularity and graininess. B indicates a hypothetical foveal cone distribution (a center-to-center distance of 1.5μ is assumed). The distance between X and Y represents the best recent estimate (see below) of the average amplitude of physiological nystagmus. X and Y mark the limits of tremor for cone *a*. Thus, the diffraction image of the geometrical border MN is spread over the cones with the graded illuminance of curve A. Tremor XY will, by shifting the cones with respect to the diffraction image, subject them to different rates of change of illuminance. For example, if the bank of cones shown in Figure 3 moves to the right from the position shown, cone *a* will receive the greatest rate of change of illuminance since it lies on the steepest portion of curve A. Cone *b*, transecting a less steep portion of the illuminance distribution, will be subjected to a less rapid stimulus onset, and consequently produce a somewhat lower neural discharge frequency. From an inspection of Figure 3, it can be seen that the middle cones (*a*, *b*, $-b$) will receive the greatest over-all illuminance changes, with cone *a*, on the average, receiving the largest rate of illuminance change. Considering the complete cone bank B, the differences in both amplitude and rate of illuminance change arising from ocular tremor will create a gradient of discharge frequency over the receptors which is peaked in the center.

At the retinal level, then, the Marshall-Talbot theory provides one mechanism for the neural peaking of borders. Before passing on to higher synaptic levels, it should be noted that, unlike the Weymouth theory, receptor size plays an important role here. Cone size is one factor determining the rate of illuminance change. As Marshall and Talbot (**53**) put it: "Smaller receptors would be useless, because though traversing the optical gradient oftener, they would gather proportionally less brightness differential. The limiting retinal factor in acuity seems to be

the relation of receptor width to the highest optical *gradient* in a moving pattern, rather than the average static differential illumination on one cone, compared with its neighbors."

The diffraction image, peaked at the retinal level by differential magnitudes and rates of illuminance change owing to the nature of the illuminance distribution and nystagmus motion, is passed on to the lateral geniculate. By referring to Figure 2, it can be seen that multiplication of pathway and reciprocal overlap will provide further peaking for submaximal reactions. If the system is not saturated the greatest reaction density will occur at the center of the reacting group, for it is here that both spatial and temporal summation will be greatest owing to the peaked character of the in-coming discharge frequency pattern. Pathway multiplication and overlap serve to refine the mosaic "in proportion to the sharper gradients and peaks produced, as sand forms sharper peaks than bricks" (**53**).

To complete the picture, we must consider yet another neurophysiological mechanism: the neural recovery cycle. Observations on the recovery cycle at the lateral geniculate of the cat show that the second of two stimuli of weak or moderate intensity "produces an enhanced postsynaptic spike, during the first 10 to 30 msec. . . . This mixture of supernormality, facilitation, recruitment, or summation is succeeded by a longer period of depression or subnormality" (**53**). The recovery cycle functions "to break up neural activity into temporally discontinuous transmission," and the summation possible during the supernormal period will serve to amplify the gradients oscillating at the lateral geniculate level. Utilizing the observations of Adler and Fliegelman (**2**) on physiological nystagmus, Marshall and Talbot (**53**) state that the periods and amplitudes of this action "yield velocities of transit across the mosaic which stimulate the receptors discontinuously, at intervals important in the neural recovery cycle."

To clarify the manner in which the recovery cycle at the lateral geniculate might serve to emphasize intensity differentials at borders, consider what might possibly occur as cone bank B (Fig. 4) oscillates back and forth. With the bank at the extreme right position (cone a at Y), cones b, and c have just recently given on-responses while cones $-b$ and $-c$ have responded somewhat earlier. As bank B begins to shift to the left on the second half of the cycle, the geniculate correlates of cones b, and c will be responding to the onset of illuminance on cones b and c in the supernormal portion of their recovery cycle. Therefore, early in the second half-cycle of nystagmus, the geniculate correlates will show summation (peaking) according to the Marshall-Talbot scheme. By the time cones $-b$ and $-c$ reach the steep portion of A their geniculate correlates will be in the subnormal phase so that the activity arising from these cones undergoes suppression at the geniculate. On other cycles, the geniculate correlates of $-b$ and $-c$ will be amplified while b and c are suppressed. The correlate of a will tend to be somewhat suppressed compared to (b, c) and ($-b$, $-c$) since recruitment at the geniculate is inversely proportional to the strength of the conditioning stimulus, and a receives, on the *average*, a stronger stimulus on its conditioning half-cycle than b, c, $-b$, or $-c$. This means that the gradients of activity at the borders of bar C will be sharpened and acuity enhanced.

It has already been indicated how

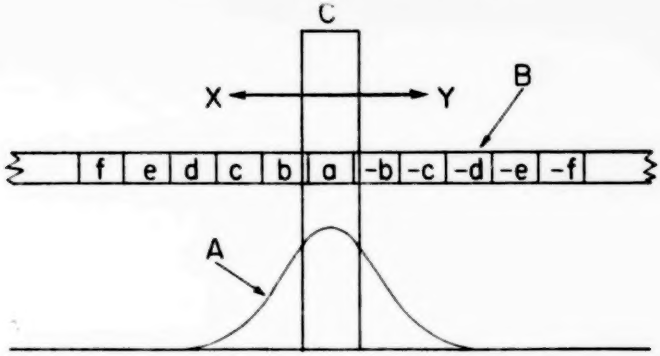

FIG. 4. THE DISTRIBUTION OF ILLUMINANCE (A) ON THE RETINA FOR A BRIGHT BAR (C) ON A DARK FIELD. B IS A BANK OF CONES, XY THE LIMITS OF PHYSIOLOGICAL NYSTAGMUS

the mechanism of pathway multiplication and overlap in combination with a fluttering neural "image" would tend to produce ever sharpening gradients at each synaptic level for lines and edges. The cortical mosaic is considered to be of a much finer grain than the retinal (see description of pathway multiplication above): " . . . quantitatively the unit paths near central vision should now be conceived, not as lines, but as expanding cylinders whose ends bear an area ratio of 1:10,000, and a cellular ratio of perhaps 1:100" (**53**). Visual resolution may be explained by differential receptor stimulation produced by ocular tremor "scanning" which ultimately gives rise to peaks of activity in the finer cortical mosaic. Since cortical excitation peaks occur upon a greatly expanded scale, and after successive refinement and sharpening at other synaptic levels, the accuracy of spatial localization will be more precise than that afforded by the coarser retinal mosaic. Reciprocal overlap " . . . while broadening the base of a local reaction, provides for small shifts of reaction peaks, if the system is not saturated" (**53**). Thus, stimuli having a spatial separation much smaller than the center-to-center distance between adjacent foveal cones will be resolved; e.g., the center-to-center distance of cones at the fovea centralis is approximately 20″ while vernier acuity is of the order of 2″.

AN EVALUATION OF FACTORS INVOLVED IN THE MARSHALL-TALBOT THEORY

Since various aspects of the Marshall-Talbot theory have been utilized to account for other visual phenomena besides acuity (see **37**, **44**, **62**, **63**), and because it is the most comprehensive and sophisticated acuity theory to date, an evaluation of the major physiological mechanisms upon which the theory rests will be attempted.

Physiological Nystagmus and Acuity

The small-amplitude, high-frequency, involuntary flutter of the eyes during "steady" fixation plays an important role in dynamic theories and particularly in the Marshall-Talbot theory. Recent evidence on the characteristics of this motion will be described, and relevant aspects of the theory will be evaluated in the light of this material. Experiments attempting to trace the influence of

fixation flutter upon acuity will be discussed.

Adler and Fliegelman (**2**) describe physiological nystagmus as having a mean amplitude of 2′ 14″ and a frequency of 50 to 100 per sec. It is these data which are used by both Marshall and Talbot (**53**) and Jones and Higgins (**44**) in their theoretical treatments. Ratliff and Riggs (**66**) recalculated the data presented by Adler and Fliegelman and arrived at a value of approximately 1′ for mean tremor amplitude. In their investigation, Ratliff and Riggs (**66**) found the median tremor to be 17.5″, with a range from just perceptible movements to almost 2′ of arc. Tremors greater than 1′ were rare. The frequency ranged from 30 to 70 cycles per sec. Other measurements of physiological nystagmus by Barlow (**5**) and Ditchburn and Ginsborg (**18**) are in good agreement with these values. However, another recent investigation by Higgins and Stultz (**43**) yielded a median tremor amplitude of 1.0′, with a mean at 1.2′. In terms of Polyak's histological description of the fovea centralis (see above), the difference between the results of Ratliff and Riggs (**66**) and Higgins and Stultz (**44**) is roughly equivalent to a flutter motion across one cone versus motion across four cones. In Figures 3 and 4 we have assumed a tremor motion across 4 cones, for illustrative purposes, in accordance with the findings of Higgins and Stultz. This may or may not be an overestimate, but in any case, the combination of tremor, drift, and flick movements make it appear "unlikely that any point of an image can remain on a single receptor for more than a few hundredths of a sec." (**66**). There do not seem to be any differences in the amplitude or frequency of tremor under changes in test-object shape (**66**), visibility, or monocular versus binocular fixation (**43**). This squares with Walls' (**79**) suggestion that the tremor is "due to proprioceptive feedback from the muscles, going no higher than the cranial-nerve nuclei themselves." Summarizing: Investigations show fairly good agreement on the frequency of ocular tremor, but the amplitude is certainly smaller than the 2′ 14″ estimate of Adler and Fliegelman (**2**). Ratliff and Riggs (**66**), using a contact lens arrangement, obtained amplitude values less than one-third as large as those found by Higgins and Stultz (**43**) who directly photographed the excursions of a small scleric blood vessel. Such measurements are technically difficult and the source of the experimental differences may lie in the type of recording device utilized.

The effect of an over-all smaller tremor amplitude upon the Marshall-Talbot theory is difficult to estimate. Barlow (**5**) feels that his small amplitude results are evidence against the theory, while Ditchburn and Ginsborg (**18**), with comparable findings, regard them as compatible with (although not directly supporting) the Marshall-Talbot theory. These latter investigators also note that the tremor is highly irregular; Marshall and Talbot assume a statistical distribution of the amplitudes. We have already explained how they also consider the tremor characteristics, as described by Adler and Fliegelman (**2**), to be rather precisely related to the neural recovery cycle at the lateral geniculate. However, Marshall and Talbot do not work out any explicit, quantitative relations here, and in another source (**52**) describe the relationship in a way which seems to involve other eye movements besides tremor. On the face of it, the posited mode of operation of this mechanism seems highly im-

probable, for a *statistical occurrence* of both horizontal and vertical components of tremor imposed upon drifting movements can scarcely mesh with the phases of the recovery cycle over the area thus involved. Before the various movements could "average up" a sharpened set of gradients *with the aid of recovery cycles*, a large saccadic flick would move the crucial portions of the diffraction image to a new portion of the cone mosaic.

But ocular tremor is also made to serve three other purposes: (*a*) The refinement and peaking of gradients brought about by physiological nystagmus interacting with reciprocal overlap at higher levels (independently of recovery cycle dynamics) would presumably still be operative for tremor of 1′ amplitude. (*b*) Such image flutter could still average up a more precise "retinal mean local sign" over time. (*c*) Ocular tremor is also assigned the role of converting the spatial illuminance gradients of the diffraction image into temporal illuminance gradients on the receptors. A fluttering motion would provide rapid rate changes in stimulus strength, thus producing high neural outputs and, as a result, better acuity.

There has been no direct experimental test of *a*, but *b* and *c* have begun to be investigated. Riggs and Ratliff (**68**) describe experiments that appear to confirm the role of some sort of mean local sign in binocular vision. Stereoscopic acuity measures reveal that a difference of 2″ of arc between the images on the eyes can be utilized as an accurate depth cue, yet median tremor according to these investigators is some 17.5″ and is not binocularly coordinated. Apparently the conception of corresponding retinal points must be revised to include processes of temporal integration producing a mean retinal location. Riggs and Ratliff (**68**) conclude "that both spatial and temporal patterns of impulses from the two eyes must somehow be combined centrally."

O'Brien and O'Brien (**59**) describe an experiment which purports to eliminate tremor as a factor in visual acuity. However, the cogency of the results has been justifiably questioned by Teuber and Bender (**75**). In a later paper, O'Brien (**57**) is of the opinion that final judgment on the tremor-acuity relationship should be deferred.

That the edges are more important determinants of acuity than the area of the test-object has been revealed by a number of investigators (**10**, **21**, **47**). The relation of this fact to tremor-scanning of gradients is obvious.

Ratliff (**65**) tested monocular acuity during a test-object exposure of 75 msec. in order to ascertain the effect of eye movements upon acuity. Drift movements during exposure of the test-object were found to be detrimental to acuity, and incorrect responses were significantly related to larger amplitudes of tremor. One might expect drift to decrease acuity by changing the location of the scanning operation of tremor, but the fact that larger amounts of tremor are associated with poorer acuity does not confirm the "dynamic" theories. This result is compatible with static, intensity-discrimination theories in which acuity arises from the simultaneous differential responses of a set of adjacent receptors. But that the functionally immobile eye is not only not the most efficient condition for seeing, but is in general quite a poor one, has been recently demonstrated (**69**). It was not actually necessary to immobilize the eye. Instead, an optical system was de-

vised which provided for displacement of a vertical dark-line test-object in correspondence with any horizontal eye movement, so that the diffraction image would always lie upon the same retinal receptors. Under these "compensated" conditions, fine test-objects disappeared after a few seconds and failed to reappear. With a viewing time of one minute, maintenance of resolution was shown to benefit when eye movements had their normal effect, and was especially aided by "exaggerated" (the amplitude of image motion due to eye movements was doubled by altering the optical system) eye movements. When short exposure times comparable to Ratliff's (**65**) value of 75 msec. were used results similar to his were obtained. Here, the "compensated" (no image movement) condition yielded the best visual performance. With somewhat longer exposure times, the normal and "exaggerated" conditions began to excel, and it is suggested that these conditions operate to "shift the acuity task from one set of receptors to another in rapid succession so that not all of the receptors at any one time have achieved a stationary state . . . eye movements are bad for acuity but good for overcoming the loss of vision due to *uniform* stimulation of the retinal receptors" (**69**, italics mine).

The investigation described above (**69**) is in many respects crucial to the question of the role of ocular tremor in acuity. The finding that with short exposures the "compensated" image was best for acuity is counter to the Marshall-Talbot theory. However, the rapid fading of fine lines under "compensation" with longer viewing time seems to indicate that once the retinal "on-response" (**32**) to the presentation of the test-object is over it will disappear unless similar responses are continuously evoked. Under normal viewing conditions, the fact that the "on" of *short exposure*, when coupled with the "on-responses" associated with diffraction *image movement* (due to tremor), should give rise to poorer acuity might be attributable to some "smearing" or confounding of the spatio-temporal relationships by the two simultaneous on-processes. Eye movements certainly produce superior acuity performances when allowed to operate over an appreciable period of time. Therefore, it seems quite possible that high acuity is normally *obtained and maintained* by "on-off responses," which during prolonged viewing are continuously evoked by means of ocular tremor. Under normal viewing conditions, high short-exposure acuity might be obtained from the "on-response" of stimulus onset and fail to be maintained by tremor due to "smearing" as explained above. This view of the processes producing acuity differs from that given by Riggs *et al.* (**69**) quoted above. We consider eye movements to be always an aid to acuity (since they provide the necessary temporal illuminance gradients) unless simultaneously confounded with other "on-response" producing effects (e.g., flash exposure of the test-object). By "on-response" is meant the on-component of "on-off" optic nerve fiber activity (**32**). This will be given further consideration in a later section.

It is possible to test the importance of "on-off responses" for acuity in a precisely controlled way by using a flickering test-object under "compensated" conditions. With such an arrangement we would predict that a fine-line test-object would not disappear. This is just what occurs,[7] but there are complications. Very

[7] Tom N. Cornsweet. Personal communication, Feb. 7, 1955.

slow flicker in the neighborhood of one cycle per sec. permits continued seeing of the test-object with no fading out, but there is a rapid decrement in the percentage of time for which the test-object is seen when flicker frequency is greater than about 1.5 cps. Electrophysiological studies on the cat indicate that the channel capacity for "on-off responses" in the retinal ganglia (**19**) and evoked cortical potentials (**56**) is much higher than 1 cps, and approximates the flicker fusion frequency. If resolution is related to the activation of "on-off" mechanisms (as we have suggested) then it is difficult to account for an optimum seeing condition around 1 cps with "compensation" and marked deterioration with higher frequencies which are well within the capacities of "on-off" mechanisms. Perhaps the supernormal period of the lateral geniculate is the important factor here.

Another piece of experimental evidence seems to militate against this view. Senders (**71**) found that when a flicker light source was used the test-object luminance required for threshold visual resolution to occur was less than would be required by the Talbot-Plateau law. Since "compensated" conditions were not used normal eye tremor presumably occurred, and flicker "on-off respones" plus tremor-induced "on-off responses" should have been confounded according to the view presented above. It is interesting to note that Senders' (**71**) enhancement of resolution occurs at flicker rates where Cornsweet, using the "compensated" condition, finds a decrement.

Evidence on Reciprocal Overlap and Multiplication of Pathway

The role of reciprocal overlap and multiplication of pathway in providing the anatomical basis for visual acuity is central to the Marshall-Talbot theory. They cite evidence of overlap in the lateral geniculate of the cat and pathway multiplication in the monkey (see above).

With respect to the possibility of overlap, the situation at the fovea is described by Polyak (**64**) as follows:

> While it is possible and even probable that under special dynamical conditions the pure cone system [i.e., single cone—single midget bipolar—single midget ganglion, the "private line" system characteristic of the central region of the fovea] functions independently, it is likely that the same system, even in the foveal center, is rarely activated without some participation of the diffuse neurons, such as brush and flat bipolars. This widening of the "roadbed" for the centripetal impulses inevitably blurs, to some degree, the distinctness of the photoreceptor processes, and in this way suppresses somewhat the effect of the clear-cut barriers separating the individual cones from one another (**64**, p. 431).

However, it should be noted that Polyak claims the midget synapses themselves do not overlap reciprocally.

We have already distinguished between sensitivity and acuity, the one referring to luminance threshold phenomena, the other to resolution of a test-object. Resolution depends upon, among other things, the diameter and spacing of the percipient elements. But also involved is the fact that beyond a small area at the center of the fovea groups of visual receptors tend to converge upon the same bipolar, and many bipolars funnel onto the same ganglion cell. The rods especially converge in this fashion, many of them using the same neural pathway. Where this mode of connection occurs it effects a coarser "neural grain" and thus reduces resolving power, but it enhances spatial summation which increases sensitivity (**32**).

Whatever the situation is at the retina, the great *reduction* in path

number at the optic nerve must be considered. In man, there are some 125 million rods and 6 million cones in the retina. But there are at most only one million fibers in the optic nerve, and only some 80 per cent of these terminate in the lateral geniculate. Any overlap or increase in "grain" which might occur in the retina will be undone (at least in any spatial sense) by the great degree of funneling at the optic nerve.

For the primary visual pathway, the lateral geniculate nucleus is the only relay center between the optic nerve and the cortex. It is a laminated structure, and Walls (**80**) describes its topological relations to the retinae as follows:

> The problem of why the LGN [lateral geniculate nucleus] should ever be stratified has seemed a mystery for half a century. In this time it has been brilliantly shown: first that each LGN embodies an isomorph of one half of the binocular visual field and the adjoining uniocular temporal crescent, with each peripheral or central quadrant of each of two hemiretinas projecting to a specific segment of the nucleus; second, that in the strongly binocular carnivores and primates certain laminae receive only from the contralateral retina and the others from the ipsilateral; and third, that physiologically correspondent spots in homonymous hemiretinas are represented by cells scattered along a single "line" running through *all* of the gray laminae like one of the toothpicks in a club sandwich. . . . But this knowledge helps us not one jot or tittle to see why the LGN should have lamination per se—not, if no cell in any lamina communicates within the nucleus with any cell in any other lamina. No matter how closely the system may bring binocularly correspondent paths into approximation, if it is *not* for the purpose of enabling them to "fuse" in the LGN, then the approximation is senseless; and, the lamination that brings it about continues to seem senseless.

Thus, there appears to be no interlaminary communication in the LGN, i.e., no evidence for binocular interaction here.[8] This is given full cognizance by Marshall and Talbot. In fact, using the cat, they were among the first to demonstrate it (**73**). There has been further confirmation by Marshall (**51**), although recently Bishop and Davis (**11**) report electrophysiological evidence of bilateral interaction in the cat LGN. In any event, this particular aspect of the LGN is not critical for the Marshall-Talbot theory one way or the other.

Marshall and Talbot do utilize the observations of Glees (**23**) on the cat LGN. For both the cat and the rabbit (**24**) Glees found multiplication of pathway and extensive reciprocal overlap *within* each cell lamina. Also cited by Marshall and Talbot is the path-multiplication seen in the Macaque monkey LGN by Glees and Clark (**25**), although there was *no evidence of reciprocal overlap*. Here, each optic fiber ended in a spray of 5 or 6 branches. The end boutons contacted the cell bodies and not the dendrites, and in all cases only *one* bouton was related to one geniculate cell. Clark's (**15**) comment on the probable function of these LGN structures is thought-provoking:

> Clearly, such an arrangement [in the monkey LGN] would provide for the utmost precision in the recording at the geniculate level of a retinal image. On the other hand, the profusion of synaptic contacts in the geniculate of the cat, together with the overlap of different optic fibers, would presumably provide an anatomical basis for a high degree of sensitivity even in low intensities of illumination.

It was noted above that funneling arrangements in the retina enhance sensitivity at the expense of acuity. Clark extends the same function to the mechanism of reciprocal overlap. The overlap, characteristic of the nocturnal animal's LGN, is a struc-

[8] However, Harman (**31**) refers to unpublished work by Fox which indicates evidence for interlaminar overlap in monkey Golgi preparations.

ture for increasing light sensitivity, while the lineation in the LGN of sharp-sighted, diurnal animals maintains optic nerve differentials for transmission to the cortex.

For the human LGN only cell-count data are available, but nonetheless Walls (**80**) concludes:

> . . . in the human LGN, each incoming optic-tract fiber synapses with one, and only one, outgoing optic-radiation fiber. There is no funneling; and there is probably neither path-multiplication nor reciprocal overlap. Overlap could theoretically exist and the numbers of incoming and outgoing fibers could still agree exactly, but there would then be an enormous difference between Man and monkey.

Further, Walls (**80**) doubts the generality of path-multiplication in the monkey LGN observed by Glées and Clark (**25**), and is of the opinion that the typical relationship here between tract fibers and radiation fibers is one to one. On the whole, when the Marshall-Talbot conception of overlap and path-multiplication at the LGN *as an acuity mechanism* is confronted with the known data it fares rather poorly.

The situation at the cortex squares with the theoretical requirements more satisfactorily. Mapping of the visual cortex (Brodmann's Area 17) shows that the fovea has a greatly expanded representation when compared with the visual periphery. Talbot and Marshall (**74**), working on the Rhesus monkey, found that 2′ of visual arc at the fovea (or 9μ) project to 1 mm. of cortex, while just beyond the fovea the relationship was approximately 18′ per mm. of cortex. Clark (**14**) has shown that one activated spot on the cortex (fired by one geniculostriate fiber) could spread its influence over a radius of some 5 mm. by means of short association fibers. This might effect a functional multiplication and overlap in the cortex, despite a paucity of terminal axonal arborization in geniculostriate fibers.

Retinal Mechanisms

It is perhaps surprising that Marshall and Talbot make scant use of Hartline's (**32**) work on the vertebrate retina. By isolating and recording from single optic nerve fibers in the bullfrog, Hartline was able to distinguish roughly three types of fibers. "On" fibers numbered some 20 per cent and responded to illuminance onset (or sharp increase) by giving a short high-frequency burst of impulses followed by a slower steady rate of discharge which was maintained as long as the stimulus continued. About 30 per cent of the fibers were of the "off" type and discharged only when the stimulus ceased to act or was suddenly decreased to a less intense value. Some 50 per cent of the fibers, designated the "on-off" type, gave bursts of high-frequency impulses only at the onset and termination (or sudden increase or decrease) of stimulation and remained silent when stimulus values were not changing.

Marshall and Talbot hint that the "on" fibers may serve to evaluate field brilliance (but see **81**), while "on-off" fibers provide the basis for more epicritic functions. Physiological nystagmus would be a very effective mechanism for generating "on-off" responses since the receptors are subjected to rapidly changing illuminance values as they sweep back and forth over the contours of the diffraction image, going in and out of action some 50 times per second. Granit has suggested that this continued elicitation of "on-off" responses must be instrumental in producing the accurate discrimination of contour (**27**, pp. 168–169). In any case, the vertebrate eye is well equipped to respond to changes

in illuminance. "Off" fibers are apparently rare in mammals, and mammalian "on-off" responses undergo complex transformations as the parameters of stimulation are altered. Granit states that the cone system is organized for the interpretation of *changes* in the visual field, and that the off-effect in "on-off" fibers plays a major role since it has a shorter latency than the on-effect and can be quickly inhibited (**27**, p. 167). The off-effect may be inhibited by re-illumination, the inhibition occurring with a shorter latency than the new "on"-response (**27**, pp. 90–92). Further experimental work by Granit (**29**) has shown that the on- and off-components of "on-off" elements are mutually inhibitory. Considered along with ocular tremor, such functional antagonism would provide a high degree of sharpening of image borders.

Hartline found that it was possible to map a receptive field on the retina for single fibers. Illumination anywhere within a tiny retinal area 1 mm. or so in diameter would produce a response in the fiber. The receptive field was of fixed type, being of the "on," "off," or "on-off" variety, and the fields of different fibers were observed to overlap one aother. A recent investigation by Kuffler (**45**, **46**) on the retina of the cat yielded a rather more complicated arrangement. Receptive fields were found to contain three zones in concentric arrangement which allowed "on," "off," or "on-off" discharges to be obtained from any *one* ganglion cell.

> There exists a central area of low threshold as tested by a small spot of light. The discharge pattern of the central region is the opposite of that found in the periphery or surround. The center may give predominantly "off," the surround "on" discharges, or the reverse. An intermediate region gives "on-off" discharges (**46**).

The type of discharge recorded at the ganglion cell depended upon a variety of factors such as background illuminance, intensity, duration, size and location of the exploring spot-stimulus within the receptive field. But perhaps of primary importance for the problem of visual acuity were the mutual modifications operating between the zones within each receptive field. Zonal interaction was studied by means of simultaneously exciting the central and marginal zones with small spot-stimuli. Stimulation of an "on" center field at the center and at the surround ("off" zone) showed that center and surround tended to suppress one another. "Off" center fields operate in an analogous fashion. Field centers tend to be dominant, but if stimulation of the surround is more intense they are inhibited. Thus, "on" center responses are modulated by inhibitory surround activity and produce a large variety of discharge patterns.

There are numerous other complex effects, not to mention the overlap of receptive fields. Our purpose is merely to indicate how zoning of the receptive field might function in acuity tasks. Kuffler (**46**) points out that with this antagonistic arrangement of central and surround zones within receptive fields, a slight shift of the eye can produce a great change in the discharge pattern. He remarks that ocular scanning movements "should be advantageous in the perception of contrast and in acuity." If we consider an "on" center field moving from left to right in Figure 3, it is clear that as the "off" surround moves into the region where the illuminance gradient is steep the center will be inhibited. As the center crosses the gradient there should be a sharp "escape" from inhibition, followed by a quick diminution of firing (suppression) owing to the reinstatement of inhibition as the left

half of the "off" surround transects the gradient. Such modes of operation in the retina would seem to be well suited for the production of sharp contours, while field overlap would refine "retinal local signs."[9]

However, certain static stimulus factors may also bring about the formation of well-defined edges. Graham and Granit (**26**) have shown that in the fovea there is inhibition of the more weakly activated area when differential stimulation is applied to two areas. Likewise, Hartline *et al.* (**34**) observed that even in the compound lateral eye of *Limulus*, "brightly illuminated areas of the eye inhibit the activity of dimly lighted regions more than the latter inhibit the activity of the former." In both cases contrast would be enhanced. The experimental data on conditions yielding best acuity performances confirm this relationship. Acuity deteriorates at surround luminances greater than the luminance of the test-object background (**13**), presumably due to suppression of the area to be resolved. Fisher (**20**) has shown that increasing the size of a dimmer surround enhances acuity.

There are still other aspects to inhibitory processes and the ways in which they may be instrumental in image resolution. That acuity should increase with light intensity is somewhat paradoxical, for under such conditions greater spatial summation should operate to increase image blurredness. Polyak (**64**, p. 430) suggests that increases in intensity not only give rise to an increased response in the cones, but also spread inhibitory influences to surrounding areas via the horizontal or amacrine cells. This collateral suppression is said to limit the areal effects of local stimulation and improve acuity. There would probably be at least two factors operative in the production of such suppression. On the one hand, there is the contrast elicited by differential stimulation, as explained above; lateral inhibition makes for higher acuity. The finding (**50**) that inhomogeneous backgrounds yield greater acuity performances than homogeneous grounds with the same test-object might be due to the generalized inhibition of spatial summation induced by inhomogeneous background stimuli. On the other hand, it has been suggested by some investigators (**16**, **49**) that light adaptation itself somehow inhibits the action of collateral channels, allowing the foveal cones to behave as single units. Willmer (**84**) has investigated the influence of luminance upon the summation area as revealed by field sizes above which subjective brightnesses remain constant. With increases in luminance above threshold levels the summation areas were found to decrease, "indicating that polysynaptic bipolar and nerve cells probably become less important; this may be one of the factors in increasing visual acuity with higher illuminances. . . ."

The fact that central flicker interaction may occur between two light patches without the disruption of fine-line discrimination in the interspace (**30**) indicates that the retina is capable of minutely differentiated responses. In this case, there appears to be interaction across an area in which collateral suppression must be manifest.

Mention should be made of the recent work of Motokawa (**55**) on the "field of retinal induction" as it bears on the problem of acuity. Using an electrostimulus method (described in **22**), fields of induction surrounding retinal images in the human eye were mapped. Field strength was strongest at figure

[9] But it is well to remember that the cat has no true fovea and is a nocturnal animal sacrificing acuity for sensitivity.

borders, and for optical illusions was distributed in ways which were related to the characteristic perceptual phenomena. The Landolt ring and vernier-type acuity test-objects were analyzed, the vernier task inducing the more marked field deformation. This result is said to account for the much higher acuity measures obtained with this figure.

Although acuity as a function of the color illumination has been explored somewhat, the factor of chromatic contrast as a basis for resolution has been almost entirely neglected and is given no place in any theory of acuity. Although it is not strictly an acuity task, discrimination of the numbers on the pseudo-isochromatic plates is entirely a wave-length discrimination, luminance contrast cues being eliminated. As Walls (**78**) puts the case:

> When at last someone determines visual acuities for colored targets on colored grounds of equal brightness, the advocates of brightness discrimination as the "sole" basis of resolution will be given new food for thought.

The aim of this section has been not primarily to argue for any specific modes of retinal interaction, but rather to indicate certain fairly complex mechanisms which might play important roles in resolution. It is felt that the Marshall-Talbot theory, as it stands, takes insufficient account of the acuity-producing actions within the retina proper. One must keep in mind that the visual system cannot resolve stimulus differentials unless they are embodied in differential retinal responses which may be channeled through the optic nerve in temporal sequences if not spatial distributions. In view of the evidence presented on the lateral geniculate nucleus (see above), it may be wise to seek for the major peaking factors at the retinal level with refinement of "grain" perhaps occurring here, at Area 17, and beyond.

REFERENCES

1. Adler, F. H. Problems in the physiology of visual acuity. *Arch. Ophthal.*, 1934, **11**, 6–19.
2. Adler, F. H., & Fliegelman, M. Influence of fixation on the visual acuity. *Arch. Ophthal.*, 1934, **12**, 475–483.
3. Adrian, E. D., & Matthews, R. The action of light on the eye. Part II. The processes involved in retinal excitation. *J. Physiol.*, 1927, **64**, 279–301.
4. Andersen, E. E., & Weymouth, F. W. Visual perception and the retinal mosaic. *Amer. J. Physiol.*, 1923, **64**, 561–591.
5. Barlow, H. B. Eye movements during fixation. *J. Physiol.*, 1952, **116**, 290–306.
6. Bartley, S. H. *Vision: a study of its basis.* New York: Van Nostrand, 1941.
7. Bartley, S. H. The psychophysiology of vision. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 921–984.
8. Berger, C. The dependency of visual acuity on illumination and its relation to the size and function of retinal units. *Amer. J. Psychol.*, 1941, **54**, 336–352.
9. Berry, R. N., Riggs, L. A., & Duncan, C. P. The relation of vernier and depth discriminations to field brightness. *J. exp. Psychol.*, 1950, **40**, 349–354.
10. Berry, R. N., Riggs, L. A., & Richards, W. The relation of vernier and depth discrimination to width of test rod. *J. exp. Psychol.*, 1950, **40**, 520–522.
11. Bishop, P. O., & Davis, R. Bilateral interaction in the lateral geniculate body. *Science*, 1953, **118**, 241–243.
12. Byram, G. M. The physical and photochemical basis of visual resolving power. Part I. The distribution of illumination in retinal images. *J. Opt. Soc. Amer.*, 1944, **34**, 571–591.
13. Chapanis, A. How we see: a summary of basic principles. In *Human factors in undersea warfare.* Washington, D. C.: National Research Council, 1949. Pp. 3–60.
14. Clark, W. E. LeGros. Observations on the association fibre system of the visual

cortex and the central representation of the retina. *J. Anat.*, 1941, **75**, 225–235.

15. Clark, W. E. LeGros. The visual centres of the brain and their connexions. *Physiol. Rev.*, 1942, **22**, 205–232.

16. Davson, H. *The physiology of the eye.* London: Churchill, 1949.

17. Dempster, W. T. Visual factors in microscopy. *J. Opt. Soc. Amer.*, 1944, **34**, 711–717.

18. Ditchburn, R. W., & Ginsborg, B. L. Involuntary eye movements during fixation. *J. Physiol.*, 1953, **119**, 1–17.

19. Enroth, Christina. The mechanism of flicker and fusion studied on single retinal elements in the dark-adapted eye of the cat. *Acta Physiol. Scand.*, 1952, **27**, Suppl. 100.

20. Fisher, B. M. The relationship of the size of the surrounding field to visual acuity in the fovea. *J. exp. Psychol.*, 1938, **23**, 215–238.

21. Fry, G. A. The relation of the configuration of a brightness border to its visibility. *J. Opt. Soc. Amer.*, 1947, **37**, 166–175.

22. Gebhard, J. W. Motokawa's studies on electric excitation of the human eye. *Psychol. Bull.*, 1953, **50**, 73–111.

23. Glees, P. The termination of the optic fibres in the lateral geniculate body of the cat. *J. Anat.*, 1941, **75**, 434–440.

24. Glees, P. The termination of optic fibres in the lateral geniculate body of the rabbit. *J. Anat.*, 1942, **76**, 313–318.

25. Glees, P., & Clark, W. E. LeGros. The termination of optic fibres in the lateral geniculate body of the monkey. *J. Anat.*, 1941, **75**, 295–308.

26. Graham, C. H., & Granit, R. Comparative studies on the peripheral and central retina. VI. Inhibition, summation, and synchronization of impulses in the retina. *Amer. J. Physiol.*, 1931, **98**, 664–673.

27. Granit, R. *Sensory mechanisms of the retina.* New York: Oxford Univer. Press, 1947.

28. Granit, R. The organization of the vertebrate retinal elements. *Ergebn. d. Physiol.*, 1950, **46**, 31–70.

29. Granit, R. Aspects of excitation and inhibition in the retina. *Proc. Roy. Soc. Lond.*, 1952, **B140**, 191–199.

30. Granit, R., & Harper, Phyllis. Comparative studies on the peripheral and central retina. II. Synaptic reactions in the eye. *Amer. J. Physiol.*, 1930, **95**, 211–228.

31. Harman, P. J. Review of G. L. Walls' *The lateral geniculate nucleus and visual histophysiology. J. comp. Neurol.*, 1954, **100**, 237–239.

32. Hartline, H. K. The neural mechanisms of vision. *Harvey Lect.*, 1941, **37**, 39–68.

33. Hartline, H. K., & Graham, C. H. Nerve impulses from single receptors in the eye. *J. cell. comp. Physiol.*, 1932, **1**, 277–295.

34. Hartline, H. K., Wagner, H. G., & MacNichol, E. F. The peripheral origin of nervous activity in the visual system. In *Cold Spr. Harb. Symp. quant. Biol.*, 1952, **17**, 125–141.

35. Hartridge, H. Visual acuity and the resolving power of the eye. *J. Physiol.*, 1922, **57**, 52–67.

36. Hartridge, H. The visual perception of fine detail. *Philos. Trans. Roy. Soc.*, 1947, **B232**, 519–671.

37. Hebb, D. O. *The organization of behavior: a neuropsychological theory.* New York: Wiley, 1949.

38. Hecht, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1928, **11**, 255–281.

39. Hecht, S. Vision: II. The nature of the photoreceptor process. In C. Murchison (Ed.) *A handbook of general experimental psychology.* Worchester, Mass.: Clark Univer. Press., 1934. Pp. 704–828.

40. Hecht, S. A theory of visual intensity discrimination. *J. gen. Physiol.*, 1935, **18**, 767–789.

41. Hecht, S., & Mintz, E. The visibility of single lines at various illuminations and the retinal basis of visual resolution. *J. gen. Physiol.*, 1939, **22**, 593–612.

42. Hecht, S., Ross, S., & Mueller, C. G. The visibility of lines and squares at high brightnesses. *J. Opt. Soc. Amer.*, 1947, **37**, 500–507.

43. Higgins, G. C., & Stultz, K. F. Frequency and amplitude of ocular tremor. *J. Opt. Soc. Amer.*, 1953, **43**, 1136–1140.

44. Jones, L. A., & Higgins, G. C. Photographic granularity and graininess. III. Some characteristics of the visual system of importance in the evaluation of graininess and granularity. *J. Opt. Soc. Amer.*, 1947, **37**, 217–263.

45. Kuffler, S. W. Neurons in the retina: organization, inhibition and excitation problems. *Cold Spr. Harb. Symp. quant. Biol.*, 1952, **17**, 281–292.

46. Kuffler, S. W. Discharge patterns and functional organization of mammalian

retina. *J. Neurophysiol.*, 1953, **16**, 37–68.

47. Lamar, E. S., Hecht, S., Schlaer, S., & Hendley, C. D. Size, shape, and contrast in detection of targets by daylight vision. I. Data and analytic description. *J. Opt. Soc. Amer.*, 1947, **37**, 531–545.

48. Linksz, A. *Physiology of the eye.* Vol. 2. *Vision.* New York: Grune & Stratton, 1952.

49. Lythgoe, R. J. The measurement of visual acuity. *Med. Res. Council, Spec. Rep. Ser.*, No. 173. London: H. M. Stationery Office, 1932.

50. McFadden, H. B. *Visual Acuity.* Duncan, Okla.: Optometric Extension Program, 1940.

51. Marshall, W. H. Excitability cycle and interaction in geniculate-striate system of cat. *J. Neurophysiol.*, 1949, **12**, 277–288.

52. Marshall, W. H., & Talbot, S. A. Relation of the excitability cycle of the geniculate-striate system to certain problems of monocular and binocular vision. *Amer. J. Physiol.*, 1941, **133**, 378–379.

53. Marshall, W. H., & Talbot, S. A. Recent evidence for neural mechanisms in vision leading to a general theory of sensory acuity. In H. Klüver (Ed.), Visual mechanisms. *Biol. Symp.*, 1942, **7**, 117–164.

54. Moon, P. *The scientific basis of illuminating engineering.* New York: McGraw-Hill, 1936.

55. Motokawa, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413–426.

56. Motokawa, K., Tukchura, S., & Ebe, M. Cortical response to intermittent stimulation with colored light in the cat. *Jap. J. Physiol.*, 1953, **3**, 191–199.

57. O'Brien, B. Vision and resolution in the central retina. *J. Opt. Soc. Amer.*, 1951, **41**, 882–894.

58. O'Brien, B., & Miller, Norma D. A study of the mechanism of visual acuity in the central retina. USAF, WADC *Tech. Rep.*, 1953, iv, 53–198.

59. O'Brien, B., & O'Brien, E. D. A test of ocular tremor and the scanning theory of visual acuity. *J. Opt. Soc. Amer.*, 1948, **38**, 1096.

60. O'Brien, B., & O'Brien, E. D. Variation in sensitivity among cones in the human fovea, and the Hecht theory of visual acuity. *J. Opt. Soc. Amer.*, 1948, **38**, 1096.

61. Ogle, K. N. On the resolving power of the human eye. *J. Opt. Soc. Amer.*, 1951, **41**, 517–520.

62. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

63. Osgood, C. E., & Heyer, A. W., Jr. A new interpretation of figural aftereffects. *Psychol. Rev.*, 1952, **59**, 98–118.

64. Polyak, S. L., *The retina.* Chicago: Univer. Chicago Press, 1941.

65. Ratliff, F. The role of physiological nystagmus in monocular acuity. *J. exp. Psychol.*, 1952, **43**, 163–172.

66. Ratliff, F., & Riggs, L. A. Involuntary motions of the eye during monocular fixation. *J. exp. Psychol.*, 1950, **40**, 687–701.

67. Riggs, L. A., Armington, J. C., & Ratliff, F. Motions of the retinal image during fixation. *J. Opt. Soc. Amer.*, 1954, **44**, 315–321.

68. Riggs, L. A., & Ratliff, F. Visual acuity and the normal tremor of the eyes. *Science*, 1951, **114**, 17–18.

69. Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. The disappearance of steadily fixated visual test objects. *J. Opt. Soc. Amer.*, 1953, **43**, 495–501.

70. Senders, Virginia L. The physiological basis of visual acuity. *Psychol. Bull.*, 1948, **45**, 465–490.

71. Senders, Virginia L. Visual resolution with periodically interrupted light. *J. exp. Psychol.*, 1949, **39**, 453–465.

72. Sloan, Louise L. Measurement of visual acuity. *Arch. Opththal.*, 1951, **45**, 704–725.

73. Talbot, S. A., & Marshall, W. H. Binocular interaction and excitability cycles in cat and monkey. *Amer. J. Physiol.*, 1941, **133**, 467–468.

74. Talbot, S. A., & Marshall, W. H. Physiological studies on neural mechanisms of visual localization and discrimination. *Amer. J. Ophthal.*, 1941, **24**, 1255–1264.

75. Teuber, H.-L., & Bender, M. B. Neuro-ophthalmology: the oculomotor system. In *Progress in neurology and psychiatry.* 1951, **4**, 148–178.

76. Tschermak-Seysenegg, A. von. *Introduction to physiological optics.* P. Boeder (Trans.) Springfield, Ill.: Charles C Thomas, 1952.

77. Walls, G. L. *The vertebrate eye and its adaptive radiation.* Bloomfield Hills, Mich.: Cranbrook Inst. of Science, 1942.

78. WALLS, G. L. Factors in human visual resolution. *J. Opt. Soc. Amer.*, 1943, **33**, 487–505.

79. WALLS, G. L. An autodemonstration of the "physiologic nystagmus." *Amer. J. Ophthal.*, 1952, **35**, 231–235.

80. WALLS, G. L., The lateral geniculate nucleus and visual histophysiology. *Univer. Calif. Publ. Physiol.*, 1953, **9**, 1–100.

81. WALLS, G. L. The filling-in process. *Amer. J. Optom.*, 1954, **31**, 329–340.

82. WEYMOUTH, F. W., ANDERSON, E. E., & AVERILL, H. L. Retinal mean local sign; a new view of the relation of the retinal mosaic to visual perception. *Amer. J. Physiol.*, 1923, **63**, 410–411.

83. WILCOX, W. W., & PURDY, D. M. Visual acuity and its physiological basis. *Brit. J. Psychol.*, 1933, **23**, 233–261.

84. WILLMER, E. N. Subjective brightness and size of field in the central fovea. *J. Physiol.*, 1954, **123**, 315–323.

Received April 1, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

THE PROBLEM OF INFERENCE FROM CURVES BASED ON GROUP DATA

W. K. ESTES
Indiana University

Papers by Sidman (**8**), Hayes (**4**), and Merrill (**6**) have raised serious questions about the validity of inferences from curves of functional relationship based on averaged data. By means of mathematical arguments and numerical illustrations, these writers have shown convincingly that " . . . given a mean curve, the form of the individual curves is not uniquely specified" (**8**, p. 268). This demonstration strikes close to home for the learning theorist. In the study of learning, we are interested in describing behavioral changes in individuals, but owing to limited control over behavioral variability must frequently depend upon averages for groups of organisms to determine functional relationships. In many areas we could scarcely remain in business if it were actually true that " . . . the mean curve does not provide the information necessary to make statements concerning the function for the individual" (**8**, p. 268). Unfortunately it is true. More accurately, it is true *if* we regard the mean curve solely as a source of inductive generalizations. This qualification suggests that possibly the fault lies, not in the averaged curves, but in our customary interpretations of them.

It is noteworthy that learning theory, even quantitative learning theory, has made rather steady progress in spite of the widespread acceptance of a false methodological assumption. Apparently inferences from averaged curves, although not necessarily correct, must in fact often be so. This being the case, researchers in learning are unlikely to give up readily the habit of computing mean curves of functional relationship. My purpose in this note is to show that we need not feel obliged to try. The group curve will remain one of our most useful devices both for summarizing information and for theoretical analysis provided only that it is handled with a modicum of tact and understanding.

The principal point to be made is that the valid treatment of averaged curves depends upon the same principles of statistical inference that have become faimilar to all of us in such cases as the analysis of variance and the chi-square test. Just as any mean score for a group of organisms could have arisen from sampling any of an infinite variety of populations of scores, so also could any given mean curve have arisen from any of an infinite variety of populations of individual curves. Therefore no "inductive" inference from mean curve to individual curve is possible, and the uncritical use of mean curves even for such purposes as determining the effect of an experimental treatment upon rate of learning or rate of extinction is attended by considerable risk. These considerations set rather severe limitations upon the use of mean curves in the study of learning. Nonetheless we can anticipate that, as so regularly turns out to be the case in scientific research, our virtue in accepting these limitations will not go unrewarded. The same type of theoretical inquiry that has led to recognition of the need for caution in handling averaged data may

be turned in a constructive direction and lead to more effective exploitation of the one defensible and important theoretical application that remains for the averaged curve—the testing of exact hypotheses about individual functions.

The first step in this direction is to recognize that the effects of averaging are not in any way capricious or unpredictable and need not be regarded as artifacts or distortions. Distortion arises only if unwarranted inferences are drawn from the mean curves. But given any specified assumption about the form of individual functions, we can proceed to deduce the characteristics to be expected of an averaged curve and then to test these predictions against obtained data. As in any problem of statistical inference, it will always be true that other assumptions might yield the same predictions. The task undertaken will be, however, to test, not the infinity of possible hypotheses, but only the one hypothesis under consideration.

In testing quantitative theories against averaged data we may be concerned either (*a*) with the form of a functional relationship or (*b*) with parameter values for the population of organisms sampled. Case *a* is illustrated by the formerly popular pastime of trying to determine "the form of the learning curve" or by the attempts to verify Hull's hypothesis that habit strength is an exponential function of number of reinforcements (**5**). Case *b* is illustrated by attempts to determine whether the slope parameter of the habit growth curve depends upon amount of reinforcement (**11**) or whether the rate and asymptote of maze learning are functions of stimulus variability (**9**).

In studies involving Case *a*, it has been customary to operate on the tacit assumption that the form of a mean curve will reflect faithfully the form of the individual curves. Since this assumption is now recognized to be unwarranted, we can no longer expect averaged data to yield any direct answer to the question, "What is the form of the individual function?" We can, however, replace this question with one which can be answered, namely, "Is the form of the mean empirical curve in accord with the assumption that the individual functions are of a given form, say $y=f(x, a, b, \cdots)$?" (In the remainder of the discussion we shall represent by f the function relating a dependent variable y to an independent variable x and parameters a, b, etc.) It becomes a specific mathematical or statistical research problem to determine for any given function f what testable predictions can be made concerning the mean curve for a group of organisms. Some preliminary considerations that may be helpful in dealing with this type of problem will be discussed below.

In studies involving Case *b* the assumption has frequently been made that if the function obtained for the individual organism is $y=f(x, a, b, \cdots)$, then the function describing the mean curve for a group of organisms should be $y=f(x, \bar{a}, \bar{b}, \cdots)$, i.e., a curve of the same form with parameters equal to the means of the corresponding individual parameters. Since the assumption is not generally true, the treatment of this case will require, first, recognizing the instances in which the assumption holds, and, second, investigating instances in which it does not hold in order to determine what information about parameter values is obtainable from the mean curve.

Classification of Functions

Relative to these problems, the

mathematical functions that we will have occasion to deal with can be classified into three types, each calling for somewhat different treatment. Let us consider briefly the problems that will arise in dealing with each of these types and illustrate some of the procedures that will prove useful in dealing with them.

Class A. *Functions unmodified by averaging.* In these cases the mean curve for the group has the form of the individual function and the parameters of the mean curve are simply the means of the corresponding individual parameters. The chief problem here is that of defining the class of functions so that we will recognize instances of it. The essential characteristics of the class will be apparent from consideration of a few examples:

1. $y=a+bx$
2. $y=a+bx+cx^2$
3. $y=a \log x$
4. $y=a \sin x+b \cos x$
5. $y=a/x.$

A numerical illustration involving one of these examples will show in a concrete way how the averaging process works out for this type of function. Suppose that we have two organisms whose behavior in a learning situation is described by the function $y=a \log x$, where a is a constant which varies in value from one organism to another, but remains fixed in value throughout learning for any one organism. Let y_1 and y_2 be response measures for the two organisms, and let the value of a be 1 for the first organism and 2 for the second. Then the course of learning for the two organisms will be described by the equations

$$y_1=\log x$$

and

$$y_2=2 \log x, \text{ respectively.}$$

Now we compute the "empirical" response measures for each organism for the first four values of the independent variable x as indicated in Table 1. Then by averaging the two response measures at each value of x, we obtain the mean "empirical" curve represented by the values in the column headed $\bar{y}$. It is clear, however, that the column of mean values also represents the values of the function $\bar{y}=1.5 \log x$. Therefore the function describing the mean curve is of the same form as the individual functions, and the parameter of the function describing the mean curve is the mean of the individual parameters.

TABLE 1

EFFECT OF AVERAGING A SIMPLE LOGARITHMIC FUNCTION

x	$\log x$	y_1	y_2	$\bar{y}$	$1.5 \log x$
1	.00	.00	.00	.00	.00
2	.30	.30	.60	.45	.45
3	.48	.48	.96	.72	.72
4	.60	.60	1.20	.90	.90

All functions belonging to this class work out similarly.[1] Stated in the simplest terms, what they all have in common is that each parameter in the function appears either alone or as a coefficient multiplying a quantity which depends only on the independent variable x. In averaging, any quantity of the latter sort factors out at each value of x and appears in the mean curve, multiplying the mean value of the parameter.

Class B. *Functions for which averaging complicates the interpretation of parameters but leaves form unchanged.* Examples of functions falling in this class[2] are

[1] See Mathematical Note 1.

1. $y = \log bx$

2. $y = \frac{1}{a} + \frac{b}{ax}$.

In the first example, we can rewrite the function in the form

$$y = \log b + \log x;$$

then it is apparent that the mean curve for a group of organisms which differ with respect to parameter b will be logarithmic in form, for the same reasons discussed in the preceding section, but will have the mean value of $\log b$ rather than $\log \bar{b}$ as the intercept constant. Thus from a mean empirical curve, we can obtain an estimate of the geometric mean of the parameter b for the organisms sampled, but no estimate of the arithmetic mean of b.

In the second example, the mean curve of y vs. $1/x$ will be linear, but the parameters of the mean curve will be the mean values of $1/a$ and b/a for the organisms sampled, so no estimate of $\bar{a}$ or $\bar{b}$ can be obtained from the averaged data.

The testing of hypotheses involving functions in this class raises no difficulties if we are interested only in the form of the function; if we wish to estimate mean parameter values or to test hypotheses involving changes in parameter values as a function of experimental treatments, then care must be taken to allow for the effects of averaging.

Class C. *Functions modified in form by averaging.* A function will fall in this class[3] if it contains any terms involving the independent variable x which will not factor out when we sum values of y over a group of organisms for a constant value of x. The most familiar example of a function belonging to this class is the "growth" curve

$$y = a + be^{-cx}$$

encountered in some guise or other in many learning theories, and given detailed discussion in Sidman's paper (**8**).

In some cases, a function belonging to this class can be moved into Class B or even Class A by means of an appropriate transformation. Take, for example, the exponential function given above. If the value of the parameter a is known for all individuals, it can be subtracted from the response measure y, leaving us with the simpler equation

$$y' = y - a = be^{-cx}.$$

The latter can be made more tractable by the logarithmic transformation

$$\log y' = \log b - cx$$

which when averaged yields

$$E(\log y') = E(\log b) - \bar{c}x,$$

where $E()$ represents the mean, or expected, value of the term in parentheses. If, then, we take logarithms (base e) of the dependent variable y' and plot the transformed variable as a function of x, both the curve for any individual and the averaged curve for a group will be linear; from the mean curve we can obtain estimates of the mean value of the parameter c and of the geometric mean of the parameter b. By means of this strategem the problem of testing the hypothesis that an exponential function holds for individual organisms has been reduced to the very simple problem of determining whether the mean curve plotted from the transformed data departs significantly from linearity. Similarly, other hypotheses that might be tested against the group data are greatly simplified. Suppose, for example, that a theoretical curve of extinction took the form of this exponential

---

[2] See Mathematical Note 2.

[3] See Mathematical Note 3.

function, with y being a response measure, x number of trials, and the asymptote a equal to zero, and that we were interested in the question whether some difference in the experimental treatments given two groups of organisms influenced rate of extinction; by means of the suggested transformation, this problem would reduce to that of testing for a difference in slope between two regression lines. A variety of transformations which may be useful in situations of this sort have been discussed by Mueller (**7**).

Even when functions in Class C cannot be moved into one of the more docile classes by any available transformation, or when for some reason transformation of the data is undesirable (as might be the case if a contemplated transformation produced heterogeneity of variances along the curve), we are not necessarily helpless. The extent to which functional form is modified by averaging will generally depend upon the dispersion of parameter values in the group of organisms sampled; thus in some cases it may be possible by studying individual curves to estimate the dispersion of parameter values in the group and determine whether the form of the mean curve can be expected to conform closely to the form of the individual functions; see, e.g., (**3**). Further, even in the case of the most refractory functions, it will usually be possible by appropriate mathematical analysis to derive the main characteristics that should be predicted for an averaged curve; an analysis of this sort for a "growth" function has been described in a recent paper (**2**).

The Role of Experimental Error

The analysis given here might be objected to on the grounds that we have considered only the effects of averaging upon data obtained from idealized organisms which behave strictly in accordance with theoretical functions. Response measures obtained from real organisms may, on the other hand, be influenced by various sources of experimental error as well as by the variables taken account of in a given theory. The objection is pertinent, but not fatal. The answer is that in testing a theoretical prediction one must make some explicit assumption about the role of experimental error in the test situation. And as in any statistical test, the validity of the conclusions will be conditional upon the degree to which such assumptions are satisfied. In some instances, it may be reasonable to assume that the contribution of experimental error is negligible; then the analyses given above will apply without modification. Frequently it will be more reasonable to operate under the assumption, routinely made in working with analysis-of-variance models, that error combines additively with treatment effects to determine the observed response measures. In this case, if we wish to test the hypothesis that a function $y=f(x, a, b, \cdots)$ holds for individuals, we will assume that the observed response measure Y for any individual is equal to the sum of y and a random variable e which represents the contribution of experimental error, i.e.,

$$Y=y+e=f(x, a, b, \cdots)+e.$$

Now if the error variable e is independent of x, and if the function f falls in our Class A, averaging of individual curves will yield a mean curve described by the function

$$\bar{Y}=\bar{y}+\bar{e}=f(x, \bar{a}, \bar{b}, \cdots)+\bar{e}.$$

If the mean value of e is zero, which will, for example, be the case whenever the distribution of errors is

normal, then the form of the mean curve will be unaffected by the error term; if the mean is not zero, then the mean function will be modified only by the addition of a constant and the plotted mean curve will be changed only by a vertical displacement. In some cases the error variable may interact with experimental variables. If the nature of the interaction can be stated explicitly, then its effects upon the averaging process can be determined by appropriate analysis. In situations where error variables and experimental variables interact in complex or unknown ways, exact tests of quantitative hypotheses will generally be impossible.

Summary

These comments are not meant to provide an exhaustive treatment of the problem of averaging. The one point I have tried to bring out clearly is that the valid interpretation of group curves[4] depends on the principles common to all problems of statistical inference. Although the form of a group mean curve does not determine the forms of the individual curves, it does provide a means of testing exact hypotheses about them. In each particular case, the procedure must be to state explicitly the hypothesis under test, and then to derive the properties that should hold for the averaged curve if the hypothesis is correct. If the predictions thus derived are in accord with data, the hypothesis remains tenable; if they are not, then the hypothesis can be rejected at some specified level of confidence. Utilized within this framework, the averaged curve can be expected to remain one of the most valuable techniques for the analysis of behavioral data, and in fact to increase progressively in value as mathematical and statistical research continue to enlarge our repertory of special devices for the handling of particular problems.

Mathematical Notes

1. A more formal criterion for class inclusion is desirable for some purposes, and may be formulated as follows.[5] Let us consider a function $y=f(x, a, b, \cdots)$. At any given value of x, we may regard y as a function of the parameters a, b, etc., and expand the function in a Taylor's series around the mean values of the parameters (**6**, **10**), obtaining the relation

$$y=f(x, \bar{a}, \bar{b}, \cdots)+(\Delta a)f_a+(\Delta b)f_b + \cdots + \frac{(\Delta a)^2}{2}f_a{}^2 + \cdots$$

where $\bar{a}+\Delta a$ is the value of the a parameter for a given organism; $f_a{}'$

[4] Throughout this discussion we have spoken in terms of mean curves obtained from groups of organisms. Similar problems arise, and similar considerations apply, however, in the case of a curve whose points represent means of repeated measures on the same organism. Parameter values associated with an individual organism may vary either systematically or randomly during the course of an experiment. In either case, we may think of each possible combination of parameter values as determining a hypothetical curve, this population of curves being sampled at each value of the independent variable. Whether the obtained mean curve should be expected to have the same form as the hypothetical individual curves will depend on the nature of the mathematical function describing the latter and on the role of experimental error, just as in the case of a group curve.

[5] A criterion proposed by Bakan (**1**), which involves expanding the function in a Maclaurin series around the point $x=0$, is not entirely satisfactory. For one thing it is frequently inapplicable. Take, for example, the functions $y=a \log x$ or $y=x^a$; in neither case are the derivatives all continuous at $x=0$, so in neither case will the series generally represent the function. The criterion suggested in the present paper will hold for all functions which can be expanded by Taylor's theorem, a class which includes all the elementary functions and, in fact, all explicit functions that the psychologist is apt to have dealings with.

represents the ith derivative of y with respect to a, evaluated at $a=\bar{a}$; and so on. When the function is averaged over a group of individuals, we obtain

$$\bar{y}=f(x,\ \bar{a},\ \bar{b},\ \cdots)+\tfrac{1}{2}\sigma_a^2 f_a^2 + \tfrac{1}{2}\sigma_b^2 f_b^2 + \cdots.$$

Our criterion for inclusion of a function in Class A may now be stated: if in the Taylor's series development, all second and higher order partial derivatives of the function with respect to parameters are zero, then the function is unmodified by averaging. Applying the criterion to $y=a \log x$, we have $f_a=\log x$; $f_a^2=0$; and therefore $\bar{y}=\bar{a} \log x$, in agreement with the conclusion reached above by a more informal route.

2. A sufficient criterion for inclusion of a function $y=f(x,\ a,\ b,\ \cdots)$ in Class B is that it does not satisfy the criterion of Class A when expanded around $\bar{a}$, $\bar{b}$, etc., but does satisfy that criterion when rewritten $y=f(x,\ u,\ v,\ \cdots)$ and expanded around $\bar{u}$, $\bar{v}$, etc. (u, v, etc. being functions of the parameters a, b, $\cdots$). In the first example under Class B above, this criterion is satisfied if we let $\log b=u$; in the second example, it is satisfied if we let $1/a=u$ and $b/a=v$.

3. If a function falls in Class C, then in the Taylor's series developments described above, some of the second or higher order derivatives will depend on x regardless of how u, v, etc. are chosen, and thus the criteria for Class A or Class B cannot be satisfied.

It will be noted that these formal criteria provide more rigorous definitions of the various classes than can be given in nonmathematical terms. However, it should be emphasized that the conclusions about inference from averaged curves that we have reached in this paper do not depend on abstruse mathematical analyses. In many practical situations, questions concerning the effects of averaging can be handled by simple numerical methods of the type illustrated in an earlier section.

REFERENCES

1. Bakan, D. A generalization of Sidman's results on group and individual functions and a criterion. *Psychol. Bull.*, 1954, **51**, 63–64.
2. Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.
3. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
4. Hayes, K. J. The backward curve: a method for the study of learning. *Psychol. Rev.*, 1953, **60**, 269–275.
5. Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.
6. Merrill, Margaret. The relationship of individual growth to average growth. *Hum. Biol.*, 1931, **3**, 37–70.
7. Mueller, C. G. Numerical transformations in the analysis of experimental data. *Psychol. Bull.*, 1949, **46**, 198–223.
8. Sidman, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49**, 263–269.
9. Wolfle, D. L. The relative efficiency of constant and varied stimulation. III. The objective extent of stimulus variation. *J. comp. Psychol.*, 1936, **22**, 375–381.
10. Woods, F. S. *Advanced calculus.* New York: Ginn, 1934.
11. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

Received February 7, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

ANALYSIS-OF-VARIANCE TESTS IN THE ANALYSIS AND COMPARISON OF CURVES[1]

DAVID A. GRANT
University of Wisconsin

The purpose of this paper is to extend methods presently available for use in the analysis and the comparison of curves obtained in psychological experiments. The reader will be assumed to have a general acquaintanceship with analysis-of-variance techniques at an elementary level, such as could be obtained from most of the better available textbooks. The present paper will deal with the analysis of curves that result when the difference between experimental treatments involves a scaled independent variable with equal steps (or equal logarithmic steps) between levels of the independent variable. Examples would be learning curves, extinction curves, dark adaptation curves, time-to-attain a level of dark adaptation as a function of the luminance of the preadaptation field, response rate as a function of the amount of reinforcement, etc. The procedures apply to most curves or sets of curves in which y, the dependent variable, is plotted as a function of x, the independent variable, where there are equal intervals between the values of x. It is assumed that there is random normal variation in the dependent variable at each data point with equal error variances. There may or may not be parameters expressing qualitative or quantitative variation of additional independent variables that are orthogonal to x.

[1] This paper was written during a research leave supported by the Graduate School Research Committee of the University of Wisconsin from funds provided by the Wisconsin Alumni Research Foundation.

Analysis of Single Trends

The first procedure, which will be briefly presented, has been described in Fisher and Yates (**6**) and Pearson and Hartley (**8**) and in other readily available sources.

As an example, consider a selected set of data from the experiment of Grant and Schiller on the generalization of the conditioned GSR to visual stimuli (**7**). In this experiment seven different groups of subjects were all conditioned to give GSR's to a 12-inch visual stimulus. The seven groups were then extinguished on 9-inch, 10-inch, 11-inch, 12-inch, 13-inch, 14-inch, and 15-inch visual stimuli, respectively. The average magnitude of GSR on the initial extinction trial, using the log-conductance transformation for 14 subjects in each group, is given in the broken curve of Figure 1.

It was expected that these data might show two general tendencies. First, although experimental steps were taken to prevent it, it was expected that there would be a tendency for the longer stimuli to produce larger GSR's. In other words, the means would tend to drift upward from left to right in Figure 1 as indicated by the line labeled "linear component." Secondly, it was expected that the data would show a generalization function which might be a symmetric, decreasing function about the 12-inch test stimulus as indicated by the curve marked "quadratic component." The obtained data seem to be a composite of both of these tendencies plus some random variation. It is obvious,

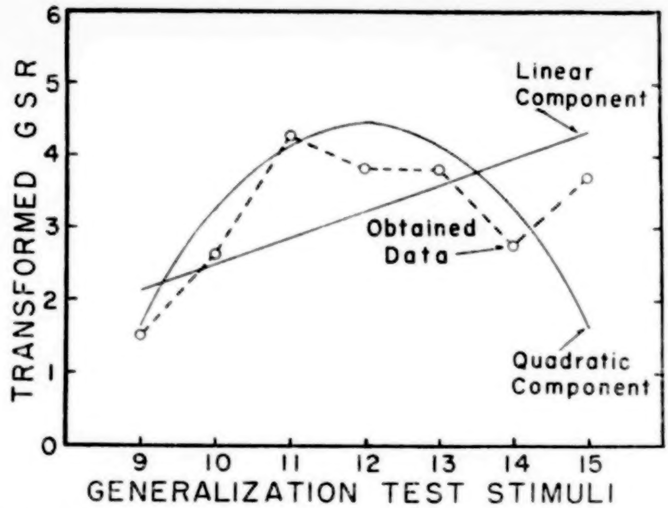

FIG. 1. FIRST GENERALIZATION TEST TRIAL DATA FROM GRANT-SCHILLER EXPERIMENT; TRANSFORMED GSR PLOTTED AGAINST VALUE OF THE TEST STIMULI

under these circumstances, that the *between groups F* is singularly unrevealing. The means for the generalization test data appear in the seond column of Table 1, and the analysis of variance is summarized in the lower part of Table 1. The *between groups* mean square is statistically significant as shown by the F of 2.23, but the conclusion permitted by this test, namely, that the group means differ significantly, has little bearing on the phenomena with which the experiment deals.

Far better tests of the specific hypotheses in which we are interested can be obtained from rather elementary mathematical and statistical considerations. In general if y, the dependent variable, is a function of x, the independent variable, then y may be expressed in the form of a power series.

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots. \quad [1]$$

Independent statistical tests of the linear, quadratic, cubic, and higher components of a curve can be obtained because the above formula may be written in the form:

$$y = A_0 + A_1\phi_1(x) + A_2\phi_2(x) + A_3\phi_3(x) + \cdots, \quad [2]$$

where $\phi_i(x)$ is the orthogonal polynomial of ith degree. Under the conditions of normal, homogeneous, random variations about the data points, and equal intervals between levels of the independent variable, A_i can be independently estimated and tested so that independent tests are available for the existence of significant linear, quadratic, cubic, etc., components of the trend. The procedure is standard (**6**, pp. 27–29; **8**, pp. 91–95).

In the Grant-Schiller experiment, the hypothesis that increasing size of stimulus is related to increasing GSR is best investigated by testing the significance of the linear component of the trend, and the hypothesis that there is a generalization gradient decreasing symmetrically

from the 12-inch test stimulus is best investigated by testing the significance of the quadratic component of the trend. The procedure for making the test is simplicity itself. First of all, the group totals are written down in order, as in column three of Table 1. There are seven data points for 6 degrees of freedom between group totals, so that the contribution of each of six orthogonal components can be tested. The values of the six orthogonal polynomials for each of the seven data points are obtained from Pearson and Hartley's table (8, p. 212). These values are given in the right-hand upper portion of Table 1, under the columns headed $\phi_{1k}, \phi_{2k}, \cdots, \phi_{6k}$. For example, the values of the fourth polynomial, ϕ_{4k}, are 3, −7, 1, 6, 1, −7, and 3, respectively. In addition, the unit coefficients have been written in column headed I. Sums of cross-products are then computed for each of the orthogonal polynomials. These are obtained by summing the products of the polynomial times the respective group totals. The sums of cross-products are labeled P_j, and they appear at the foot of the upper part of Table 1. Under column headed I is the sum of all of the group totals or the grand total $\overline{GT}$. To obtain P_1, the sum of cross-products for the 1st degree polynomial, we have: $21.34(-3)+37.01(-2)+59.87(-1)+54.09\,(0)+54.10\,(1)+39.03\,(2)+51.86(3)=89.83$. Similarly, under the column ϕ_{4k} we have $21.34(3)+37.01(-7)+59.87\,(1)+54.09\,(6)+54.10\,(1)+39.03\,(-7)+51.86\,(3)=125.83$. Also given in Table 1 is $\sum_k \phi_{jk}^2$ for each of the six sets of

TABLE 1

ANALYSIS OF GENERALIZATION CURVE FROM GRANT-SCHILLER DATA

Group	M_k	T_k	I	ϕ_{1k}	ϕ_{2k}	ϕ_{3k}	ϕ_{4k}	ϕ_{5k}	ϕ_{6k}
9	1.52	21.34	1	−3	5	−1	3	−1	1
10	2.64	37.01	1	−2	0	1	−7	4	−6
11	4.28	59.87	1	−1	−3	1	1	−5	15
12	3.86	54.09	1	0	−4	0	6	0	−20
13	3.86	54.10	1	1	−3	−1	1	5	15
14	2.79	39.03	1	2	0	−1	−7	−4	−6
15	3.70	51.86	1	3	5	1	3	1	1
$P_j=\Sigma\phi_{jk}T_k$			317.30	89.83	−192.27	34.27	125.83	−6.41	244.71
$\Sigma_k\phi_{jk}^2$				28	84	6	154	84	924

Summary of Analysis of Variance

Source of Variation	df	Sum of Squares	Mean Square	F
A. Error (Within Groups)	91	531.4938	5.8406	—
B. Between Groups	6	78.0098	13.0016	2.2261*
C. Linear	1	20.5853	20.5853	3.52
D. Quadratic	1	31.4352	31.4352	5.38*
E. Cubic	1	13.9813	13.9813	2.39
F. Quartic	1	7.3438	7.3438	1.26
G. Quintic	1	0.0349	0.0349	0.01
H. Sextic	1	4.6292	4.6292	0.79

* Significant at the 5% confidence level.

polynomial coefficients. Thus, for the first orthogonal polynomial $\sum_k \phi_{1k}^2 = (-3)^2 + (-2)^2 + (-1)^2 + (1)^2 + (2)^2 + (3)^2 = 28$. The sum of squares for each orthogonal component is then simply: $SS_{P_j} = P_j^2 / n \sum_k \phi_{jk}^2$ where n is the number of scores making up each group total. More generally, this formula can be written:

$$SS_{P_j} = \left(\sum_k T_k \phi_{jk} \right)^2 \Big/ \left(n \sum_k \phi_{jk}^2 \right), \quad [3]$$

where the subscript, j, refers to the order of the polynomial and the subscript, k, refers to the level of the independent variable.

For the linear component, then, the sum of squares will equal $(89.83)^2/(14)(28)$ or 20.5853; and for the 4th polynomial the sum of squares will be $(125.83)^2/(14)(154)$ or 7.3438. The sums of squares are entered in the summary of the analysis of variance table. A good check on the computations is that the sum of squares for all of the orthogonal components must equal the $SS_{\text{between groups}}$. In this particular instance, only the linear and quadratic sums of squares would be computed to test specific hypotheses, and the additional component sums of squares need be computed only to check the accuracy of computations.

When the mean squares resulting from these components are tested against the error variance, it is found that only the quadratic component is statistically significant, although the linear component approaches significance. It is therefore possible to conclude that there is a significant quadratic trend in the data of Figure 1, or that there is a statistically significant generalization effect. The other orthogonal components have been tested in Table 1, but, for reasons given by Duncan (**4**), it is not a particularly wise procedure to test components for which there are not specific a priori hypotheses.

The above analysis can generally be used when a different group of subjects is run under each of the levels of the independent variable. When the same subjects have been used in repeated tests under different values of the independent variable, it is necessary to subdivide the SS_{error} into orthogonal components in order to obtain appropriate error terms for testing each of the components of the over-all trend. The procedure for subdividing the interaction will be outlined as a special case of the general procedure described below for repeated measurements.

A word of caution should be inserted here with respect to logarithmic, exponential, and trigonometric functions. If there are theoretical bases for expecting such transcendental components in the data, and the data are highly reliable, all of the orthogonal polynomial components may be statistically significant, and a more definitive test may be obtained by fitting the appropriate parameters. This type of test is beyond the scope of the present paper.

Orthogonal Polynomial Analysis of Trends Based upon Repeated Measurements of the Same Ss

Alexander (**1**) has presented a very useful technique for the analysis of trends based upon repeated measurements on the same individuals. The present paper extends the procedures outlined by Alexander by further analysis of the orthogonal components of the trend and by providing for the separation of orthogonal components of differences between groups which may be permitted

by the logic of the experimental design. The notation used is adapted from Alexander (**1**).

The general functional plan of the analysis is as follows: We have a total of n subjects in p groups, and each subject has contributed a score on each of k trials or stages of the experiment, so that there will be nk scores in all. These give the "individual trends." Each group will then have a mean on each trial so that there are *pk group by trial* means which give the "group trends." There will also be a combined group mean on each of the k trials which gives the "over-all trend" of the scores. Ignoring the scaled or sequential character of the *trials* variable, the total sum of squares could be analyzed into: SS_{trials} with $k-1$ *df*, SS_{groups} with $p-1$ *df*; $SS_{\text{individuals}}$ with $n-p$ *df*; $SS_{\text{groups}\times\text{trials}}$ with $(k-1)(p-1)$ *df*; and $SS_{\text{individuals}\times\text{trials}}$ with $(k-1)(n-p)$ *df*. But *trials* is a scaled variable, and the Alexander (**1**) scheme of trend analysis subdivides SS_{trials} into $SS_{\text{over-all slope}}$ with 1 *df* and $SS_{\text{over-all deviations from linearity}}$ with $k-2$ *df*; similarly $SS_{\text{groups}\times\text{trials}}$ is broken down into $SS_{\text{between group slopes}}$ with $p-1$ *df* and $SS_{\text{group deviations from linearity}}$ with $(p-1)(k-2)$ *df*—these deal with the differences between groups in linear components and nonlinear components of trend, respectively; and the $SS_{\text{individuals}\times\text{trials}}$ is broken down into $SS_{\text{between individual slopes}}$ with $n-p$ *df* and $SS_{\text{individual deviations from estimation}}$ with $(n-p)(k-2)$ *df* which consists of the differences between individuals in linear components and nonlinear components of trend, respectively. Our method goes on to separate the $SS_{\text{over-all deviations from linearity}}$, $SS_{\text{group deviations from linearity}}$, and $SS_{\text{individual deviations from estimation}}$ into the quadratic, cubic, quartic, etc., components of the *over-all trend*, the differences *between group trends*, and the differences *between individual trends*, respectively, so that we have *between group quadratics*, *between group cubics*, etc., and *between individual quadratics*, *between individual cubics*, etc.

If the p groups form an orthogonal design, e.g., variables A and B with $a-1$ and $b-1$ *df*, respectively and an AB interaction with $(a-1)(b-1)$ *df*, then each sum of squares between groups (means, slopes, quadratics, cubics, etc.) can be separated into an A, a B, and an AB interaction component. This extension of Alexander's procedure is frequently helpful, and it too has been included in the example below.

As an example, selected data are presented in Table 2 and Figure 2 from an unpublished experiment by Grant, Kuboyama, and Patel on the influence of electric shock stimulation on the conceptual behavior of "anxious" and "nonanxious" *Ss*. In Table 2, on the left-hand side appear the perseverative error scores on the second, third, fourth, fifth, and sixth stages of the Wisconsin Card Sorting Test (WCST) as subjected to the square-root transformation (**2**). The average values of the transformed scores have been plotted for each group on the successive stages in Figure 2. In this experiment high and low anxiety groups were selected on the basis on the Taylor Anxiety Scale (**10**) and were subdivided into three subgroups each, so that the subgroups could receive 0, 2, and 12 electric shocks during the course of the experiment. The *Ss* receiving two shocks received them at the first stage of the WCST and the *Ss* receiving 12 shocks received two shocks per stage on each of the six stages of the WCST. High-anxiety groups were designated H_0, H_2, and H_{12}, respectively, and the low anxiety groups were designated L_0, L_2, and

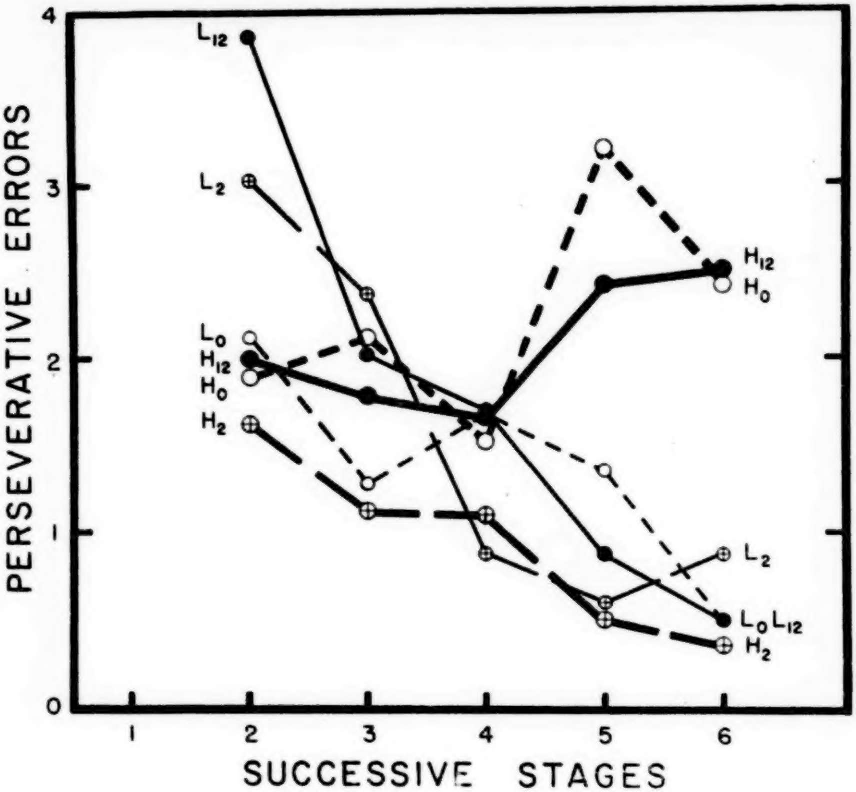

FIG. 2. TRANSFORMED PERSEVERATIVE ERROR SCORES PLOTTED AT EACH STAGE OF THE WISCONSIN CARD SORTING TEST FOR THE SIX EXPERIMENTAL GROUPS

L_{12}, respectively. Data from selected sets of four of the 17 *S*s in each original experimental group appear in Table 2.

From Figure 2 it is apparent that, as expected, two of the high-anxiety groups showed progressive deterioration in performance during the course of the WSCT. In contrast, the low-anxiety groups showed progressive improvement throughout the experiment as did also the H_2 group. Overall tests of the differences between groups and differences between stages are not particularly revealing with respect to the trends obtained. What is needed is a test which will show whether or not the improvement in the low-anxiety groups as contrasted with the failure of the high-anxiety groups to improve is statistically significant. If the low-anxiety groups improve and the high-anxiety groups deteriorate, this should result in statistically significant differences in the linear components of the trends.

The computational procedure for the orthogonal polynomial trend analysis is essentially simple and straightforward, but there are frequent opportunities for error. A number of excellent checking procedures are available, but as with all calculations, extreme care must be

TABLE 2

SELECTED TRANSFORMED PERSEVERATIVE ERRORS AND SUMS OF CROSS-PRODUCTS FOR TREND ANALYSIS

Group	Subject	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6	I	P_1	P_2	P_3	P_4
L_0	3	1.7	1	1.7	1	1	6.4	− 1.4	0.0	− 0.7	4.9
	15	3	1.7	1	1	1	7.7	− 4.7	3.3	− 0.6	− 0.8
	17	2.4	1.4	3	2.4	0	9.2	− 3.8	− 5.0	− 4.4	5.2
	21	1.4	1	1	1	0	4.4	− 2.8	− 1.2	− 1.4	− 0.6
Tot.	T_{1k}	8.5	5.1	6.7	5.4	2.0	27.7	−12.7	− 2.9	− 7.1	8.7
L_2	5	2	2.4	1	0	0	5.4	− 6.4	− 0.4	2.8	− 1.6
	16	2.8	1	1	1	1	6.8	− 3.6	3.6	− 1.8	1.8
	11A	5.3	2	0	0	1	8.3	−10.6	10.6	− 0.3	− 1.7
	18	2	4.1	1.4	1.4	1.4	10.3	− 3.9	− 1.5	4.8	−10.2
Tot.	T_{2k}	12.1	9.5	3.4	2.4	3.4	30.8	−24.5	12.3	5.5	−11.7
L_{12}	7	2.2	1.7	1.7	1.4	1	8.0	− 2.7	− 0.1	− 0.6	1.0
	21	4.4	0	1.7	1	0	7.1	− 7.8	4.4	− 6.4	10.6
	22	2	4.7	1.4	1	1	10.1	− 5.7	− 2.5	6.4	−11.4
	27	2.8	1.7	2	0	0	6.5	− 7.3	− 0.1	0.6	8.0
Tot.	T_{3k}	11.4	8.1	6.8	3.4	2.0	31.7	−23.5	1.7	0.0	8.2
H_0	24	1.4	1	0	1.4	1	4.8	− 0.4	2.4	− 1.2	− 7.2
	24A	1.7	2.4	2	3.7	2.2	12.0	2.3	− 2.3	− 2.1	− 8.5
	43	1.7	4.1	1.7	4	3.2	14.7	2.9	− 1.7	1.7	−17.3
	44	2.8	1	2.4	3.7	3.2	13.1	3.5	2.5	− 5.0	1.6
Tot.	T_{4k}	7.6	8.5	6.1	12.8	9.6	44.6	8.3	0.9	− 6.6	−31.4
H_2	12	2	1.7	2.4	1	1.4	8.5	− 1.9	− 0.7	0.8	7.0
	25	1.7	1.4	0	0	0	3.1	− 4.8	2.0	1.1	− 3.9
	32	1.4	0	1	0	0	2.4	− 2.8	0.8	− 1.4	7.4
	48	1.4	1.4	1	1	0	4.8	− 3.2	− 1.6	− 0.6	− 2.2
Tot.	T_{5k}	6.5	4.5	4.4	2.0	1.4	18.8	−12.7	0.5	− 0.1	8.3
H_{12}	32A	1.4	1	2	2.4	1	7.8	0.6	− 2.6	− 3.2	0.8
	25	1.4	2.4	1	1.4	2.8	9.0	1.8	2.6	3.4	− 5.0
	30	1	2	1.4	1	1.4	6.8	− 0.2	− 1.0	2.4	− 1.2
	40	4.2	1.7	2.2	4.9	4.7	17.7	4.2	6.8	− 5.9	− 4.3
Tot.	T_{6k}	8.0	7.1	6.6	9.7	9.9	41.3	6.4	5.8	− 3.3	− 9.7
T	T_k	54.1	42.8	34.0	35.7	28.3	194.9	−58.7	18.3	−11.6	−27.6

exercised to insure accuracy. Subscripts will be used as follows: i refers to groups, j refers to the order of the polynomial, k refers to the trial number and l refers to individuals. The total number of subjects is n; the number of subjects in the ith group is n_i; there are K trials in all, and the total number of groups is p.

The first step is to enter all of the original scores, by groups, into a table such as Table 2. In the left-

hand part of the table the scores of each subject are entered in a separate row, and space is kept for group totals. The marginal totals are then computed, the row marginals in the column headed I, and the column marginals for each group headed in the rows labeled T_{1k}, T_{2k}, etc. Thus, for the first row $I = 1.7 + 1 + 1.7 + 1 + 1 = 6.4$. In the first column of the first group the $T_{1k} = T_{11} = 1.7 + 3 + 2.4 + 1.4 = 8.5$. For each group, the sum of the I's and the sum of the T_{ik}'s must be equal, which serves as a check on the computation of the marginal totals. At the foot of the table, the totals for each stage, $T_{.k}$, 54.1, 42.8, · · ·, 28.3, are computed and these add up to the grand total, $\overline{GT}$, of 194.9 which must equal the sum of the group subtotals; i.e., $27.7 + 30.8 + 31.7 + 44.6 + 18.8 + 41.3 = 194.9$.

The next step is to compute and check the P's. First the table of the orthogonal polynomials, ϕ_{jk}, for the number of trials or stages is found in Fisher and Yates (**6**) or Pearson and Hartley (**8**), and these are given in Table 3 for the present example. To obtain the P's for each subject, the following formula is used: $P_j = \sum_k y_k\phi_{jk}$. For example, for the first subject $P_1 = 1.7(-2) + 1(-1) + 1.7(0) + 1(1) + 1(2) = -1.4$, $P_2 = 1.7(2) + 1(-1) + 1.7(-2) + 1(-1) + 1(2) = 0.0$, $P_3 = 1.7(-1) + 1(2) + 1.7(0) + 1(-2) + 1(1) = -0.7$, and $P_4 = 1.7(1) + 1(-4) + 1.7(6) + 1(-4) + 1(1) = 4.9$. The best procedure is to compute all the P_1's and then all the P_2's, etc. As a check for each group, the sum of the individual P_j's for that group should be equal to the sum of the products of the stage totals, T_{ik}, for the group, times the values of the orthogonal polynomials for each stage; i.e., for the jth polynomial, $\sum_k T_{ik}\phi_{jk} = \sum_l P_{jl}$. For example, the P_1's for the first group, -1.4 -4.7 -3.8 and -2.8 add up to -12.7. Also $8.5(-2) + 5.1(-1) + 6.7(0) + 5.4(1) + 2.0(2) = -12.7$, which serves as a check on the accuracy of the P_1's for the first group. A similar check must be made with each polynomial in each group and with the over-all stage totals at the foot of Table 2.

TABLE 3

VALUES OF THE ORTHOGONAL POLYNOMIALS FOR ANALYSIS OF THE PERSEVERATIVE ERROR TRENDS

ϕ_{jk}	Stage					$\Sigma_k\phi_{jk}^2$
	2	3	4	5	6	
ϕ_{1k}	−2	−1	0	1	2	10
ϕ_{2k}	2	−1	−2	−1	2	14
ϕ_{3k}	−1	2	0	−2	1	10
ϕ_{4k}	1	−4	6	−4	1	70

The next step is to obtain the quantities specified by the equations in Table 4. Table 4 consists of three columns, U, V, and W which deal with the *trials by individuals* measures, *trials by groups* measures, and *over-all trials* measures, respectively. Associated with each entry in Table 4 is the corresponding *df*. The computations indicated in Table 4 have been carried out in Table 5 so that they can be identified from the corresponding entries in Table 2. If the work has been done correctly, the first entry in each column is the sum of the remaining entries, and this is an important check.

The only step remaining before the final computations of the sums of squares for the summary of the analysis of variance is to obtain subtotals of the groups on I, P_1, P_2, P_3, and P_4, for the three levels of shock, 0, 2 and 12, and the two levels of anxiety, L and H. This has been done in Table 6 where the entries are simply subtotals from the right-hand sides of Table 2.

TABLE 4

COMPUTATION OF THE SUMS OF SQUARES OF THE ORTHOGONAL COMPONENTS OF TREND

	U Trials by Individuals	V Trials by Groups	W Trials
Y	$\sum_{i,k,l} Y_{ikl}^2$ $df=nK$	$\sum_{i,k} \frac{T_{ik}^2}{n_i}$ $df=pK$	$\frac{1}{n}\sum_{k}\left(\sum_{i} T_{ik}\right)^2$ $df=K$
I	$\frac{1}{K}\sum_{l} I_l^2$ $df=n$	$\sum_{i} \frac{\left(\sum_{k} T_{ik}\right)^2}{Kn_i}$ $df=p$	$\frac{GT^2}{nK}$ $df=1$
P_1	$\frac{\sum_{l} P_{1l}^2}{\sum_{k} \phi_{1k}^2}$ $df=n$	$\frac{\sum_{i} P_{i1}^2}{n_i \sum_{k} \phi_{1k}^2}$ $df=p$	$\frac{\left(\sum_{l} P_{1l}\right)^2}{n \sum_{k} \phi_{1k}^2}$ $df=1$
P_2	$\frac{\sum_{l} P_{2l}^2}{\sum_{k} \phi_{2k}^2}$ $df=n$	$\frac{\sum_{i} P_{i2}^2}{n_i \sum_{k} \phi_{2k}^2}$ $df=p$	$\frac{\left(\sum_{l} P_{2l}\right)^2}{n \sum_{k} \phi_{2k}^2}$ $df=1$
P_3	$\frac{\sum_{l} P_{3l}^2}{\sum_{k} \phi_{3k}^2}$ $df=n$	$\frac{\sum_{i} P_{i3}^2}{n_i \sum_{k} \phi_{3k}^2}$ $df=p$	$\frac{\left(\sum_{l} P_{3l}\right)^2}{n \sum_{k} \phi_{3k}^2}$ $df=1$
P_4	$\frac{\sum_{l} P_{4l}^2}{\sum_{k} \phi_{4k}^2}$ $df=n$	$\frac{\sum_{i} P_{i4}^2}{n_i \sum_{k} \phi_{4k}^2}$ $df=p$	$\frac{\left(\sum_{l} P_{4l}\right)^2}{n \sum_{k} \phi_{4k}^2}$ $df=1$

The final step remains of entering the sums of squares into the analysis of variance summary table as has been done in Table 7. All the sums of squares in Table 7, except those having to do with the subdivision of components of the sums of squares related to *anxiety*, *shock*, and their *interation*, can be computed directly from the sums of squares entries in Table 5 as indicated in the column of Table 7 headed "Computation." The number of *df* for each row of Table 7 is obtained by applying the "Computation" formulae to the *df* entries in Table 5. Then the subdivision of the components of variation between group means and be-

TABLE 5

COMPUTED VALUES OF THE TABLE 4 COMPONENTS OF THE TREND ANALYSIS OF PERSEVERATIVE ERRORS

	U: Trials×Individuals	V: Trials×Groups	W: Trials
Y	$U_Y=1.7^2+1^2+\cdots+4.9^2+4.7^2=482.5300$ 120 *df*	$V_Y=\frac{1}{4}(8.5^2+5.1^2+\cdots+9.7^2+9.9^2)=390.0825$ 30 *df*	$W_Y=\frac{1}{24}(54.1^2+42.8^2+\cdots+28.3^2)=332.9179$ 5 *df*
I	$U_I=\frac{1}{5}(6.4^2+7.7^2+\cdots+17.7^2)=376.0140$ 24 *df*	$V_I=\frac{1}{20}(27.7^2+30.8^2+\cdots+41.3^2)=338.4555$ 6 *df*	$W_I=\frac{(194.9)^2}{120}=316.5501$ 1 *df*
P_1	$U_{P_1}=\frac{1}{10}[(-1.4)^2+(-4.7)^2+\cdots+(4.2)^2]=47.3850$ 24 *df*	$V_{P_1}=\frac{1}{40}[(-12.7)^2+(-24.5)^2+\cdots+(6.4)^2]=39.6232$ 6 *df*	$W_{P_1}=\frac{(-58.7)^2}{240}=14.3570$ 1 *df*
P_2	$U_{P_2}=\frac{1}{14}[(0.0)^2+(3.3)^2+\cdots+(6.8)^2]=19.9521$ 24 *df*	$V_{P_2}=\frac{1}{56}[(-2.9)^2+(12.3)^2+\cdots+(5.8)^2]=3.5230$ 6 *df*	$W_{P_2}=\frac{(18.3)^2}{336}=0.9967$ 1 *df*
P_3	$U_{P_3}=\frac{1}{10}[(-0.7)^2+(-0.6)^2+\cdots+(-5.9)^2]=23.9300$ 24 *df*	$V_{P_3}=\frac{1}{40}[(-7.1)^2+(5.5)^2+\cdots+(-3.3)^2]=3.3780$ 6 *df*	$W_{P_3}=\frac{(-11.6)^2}{240}=0.5607$ 1 *df*
P_4	$U_{P_4}=\frac{1}{70}[(4.9)^2+(-0.8)^2+\cdots+(-4.3)^2]=15.2488$ 24 *df*	$V_{P_4}=\frac{1}{280}[(8.7)^2+(-11.7)^2+\cdots+(-9.7)^2]=5.1027$ 6 *df*	$W_{P_4}=\frac{(-27.6)^2}{1680}=0.4534$ 1 *df*

tween group trends can be obtained from the entries in Table 6 by the usual analysis of variance methods; i.e., $SS_q=(\sum T_q^2/Kn_q)-C$, where the subscript, q, represents the subtotal for anxiety of shock level. Thus, for example, the sum of squares between group means for anxiety is equal to $1/60(90.2^2+104.7^2)=W_I$ or 1.7521. The sum of squares between group means for shock is equal to: $1/40\ (72.3^2+49.6^2+73.0^2)-W_I$

TABLE 6

SUBTOTALS OF SUMS AND CROSS-PRODUCTS FOR COMPUTATION OF "SHOCK" AND "ANXIETY" SUMS OF SQUARES

Measure	Shock 0	Shock 2	Shock 12
I	$I_0 = 72.3$	$I_2 = 49.6$	$I_{12} = 73.0$
P_1	$P_{1.0} = -4.4$	$P_{1.2} = -37.2$	$P_{1.12} = -17.1$
P_2	$P_{2.0} = -2.0$	$P_{2.2} = 12.8$	$P_{2.12} = 7.5$
P_3	$P_{3.0} = -13.7$	$P_{3.2} = 5.4$	$P_{3.12} = -3.3$
P_4	$P_{4.0} = -22.7$	$P_{4.2} = -3.4$	$P_{4.12} = -1.5$

Measure	Anxiety L	Anxiety H
I	$I_L = 90.2$	$I_H = 104.7$
P_1	$P_{1.L} = -60.7$	$P_{1.H} = 2.0$
P_2	$P_{2.L} = 11.1$	$P_{2.H} = 7.2$
P_3	$P_{3.L} = -1.6$	$P_{3.H} = -10.0$
P_4	$P_{4.L} = 5.2$	$P_{4.H} = -32.8$

$=8.8611$, and the interaction between group means can be obtained by subtracting these two sums of squares from 21.9054, line B, in Table 7, which gives 11.2922. Similarly, the sum of squares for the linear component of the difference in trends due to anxiety is: $1/120(P_{1.L}^2+P_{1.H}^2)-W_{P_1}$. The 120 is 12, the number of subjects per anxiety level, times 10, the sum of the squared polynomial coefficients. The sum of squares for the linear component due to shock is $1/80(P_{1.0}^2+P_{1.2}^2+P_{1.12}^2)-W_{P_1}$, where the 80 is 8, the number of subjects per shock level, times 10, the sum of the squared polynomial coefficients. The sum of squares for the linear component of the interaction is the sum of squares of the linear component between group trends, line C.1 in Table 7, minus the sum of squares for the linear component due to anxiety and the sum of squares for the linear component due to shock. The actual numerical values for anxiety will be $1/120[(-60.7^2+(2.0)^2]-14.3570=16.3804$. The sum of squares for shock is $1/80[(-4.4)^2+(-37.2)^2+(-17.1)^2]-14.3570=6.8381$. The sum of squares for the linear component of interaction will be $25.2662-16.3804-6.8381=2.0477$. Similar quantities are computed for the quadratic, cubic, and quartic components of the trend as indicated in the Computation column (of sums of squares) in Table 7. The general computing formula is:

$$SS_{P_{iq}}=\frac{1}{\sum_k \phi_{jk}^2}\sum_q \frac{\left(\sum_s P_{ij}\right)^2}{n_q}, \quad \ldots [4]$$

where the subscript, s, refers to a subtotal on P_{ij} for a single level of shock or anxiety.

The mean squares of Table 7 are computed in the usual way by dividing the sums of squares by the appropriate degrees of freedom. It remains to determine the appropriate error variances for the F tests. In general, these are the between-individuals mean squares, and the row entry for each error term has been listed in the next to last column of Table 7. The between-group-means mean squares should be tested against the between-individual-means error term. The between-group-trends-linear terms and the over-all-trends-linear term should be tested against the between-individuals-linear mean square; and, in general, the quadratic terms should be tested against the between-individuals-quadratic mean square, the cubic terms should be tested against the between-individual-cubic mean square, and the quartic terms should be tested against the between-individuals-quartic mean square. Appropriate tests for over-all trend, line A of Table 7, and between-group trends, line C of Table

TABLE 7

SUMMARY OF THE ANALYSIS OF THE TRENDS OF THE PERSEVERATIVE ERROR SCORES

Source of Variation	df	Sums of Squares: Computation	Sums of Squares: Value	Mean Square	Error Term (Row)	F
A. Over-all Trend	(4)	$W_Y - W_I$	(16.3678)	(4.0920)	E	(5.37)†
1. Linear	1	W_{P_1}	14.3570	14.3570	E.1	33.30†
2. Quadratic	1	W_{P_2}	0.9957	0.9967	E.2	1.09
3. Cubic	1	W_{P_3}	0.5607	0.5607	E.3	0.49
4. Quartic	1	W_{P_4}	0.4534	0.4534	E.4	0.80
B. Between Group Means	(5)	$V_I - W_I$	(21.9054)	(4.3811)	D	(2.10)
a. Anxiety	1	$\frac{1}{60}(I_L^2 + I_H^2) - W_I$	1.7521	1.7521	D	0.84
b. Shock	2	$\frac{1}{40}(I_0^2 + I_2^2 + I_{12}^2) - W_I$	8.8611	4.4306	D	2.12
c. Interaction	2	$SS_{bgm} - (SS_A + SS_S)$	11.2922	5.6461	D	2.71
C. Between Group Trends	(20)	$V_Y - W_Y - V_I + W_I$	(35.2592)	(1.7630)	E	(2.31)
1. Linear	(5)	$V_{P_1} - W_{P_1}$	(25.2662)	(5.0532)	E.1	(11.72)†
a. Anxiety	1	$\frac{1}{120}(P_{1.L}^2 + P_{1.H}^2) - W_{P_1}$	16.3804	16.3804	E.1	37.99†
b. Shock	2	$\frac{1}{80}(P_{1.0}^2 + P_{1.2}^2 + P_{1.12}^2) - W_{P_1}$	6.8381	3.4190	E.1	7.93*
c. Interaction	2	$SS_{P_1} - (SS_{P1.A} + SS_{P1.S})$	2.0477	1.0238	E.1	2.37
2. Quadratic	(5)	$V_{P_2} - W_{P_2}$	(2.5263)	(0.5053)	E.2	(0.55)
a. Anxiety	1	$\frac{1}{168}(P_{2.L}^2 + P_{2.H}^2) - W_{P_2}$	0.0453	0.0453	E.2	0.05
b. Shock	2	$\frac{1}{112}(P_{2.0}^2 + P_{2.2}^2 + P_{2.12}^2) - W_{P_2}$	1.0041	0.5020	E.2	0.55
c. Interaction	2	$SS_{P_2} - (SS_{P2.A} + SS_{P2.S})$	1.4769	0.7384	E.2	0.81
3. Cubic	(5)	$V_{P_3} - W_{P_3}$	(2.8173)	(0.5635)	E.3	(0.49)
a. Anxiety	1	$\frac{1}{120}(P_{3.L}^2 + P_{3.H}^2) - W_{P_3}$	0.2940	0.2940	E.3	0.26
b. Shock	2	$\frac{1}{80}(P_{3.0}^2 + P_{3.2}^2 + P_{3.12}^2) - W_{P_3}$	2.2861	1.1430	E.3	1.00
c. Interaction	2	$SS_{P_3} - (SS_{P3.A} + SS_{P3.S})$	0.2372	0.1186	E.3	0.10
4. Quartic	(5)	$V_{P_4} - W_{P_4}$	(4.6493)	(0.9299)	E.4	(1.65)
a. Anxiety	1	$\frac{1}{840}(P_{4.L}^2 + P_{4.H}^2) - W_{P_4}$	0.8596	0.8596	E.4	1.52
b. Shock	2	$\frac{1}{560}(P_{4.0}^2 + P_{4.2}^2 + P_{4.12}^2) - W_{P_4}$	0.4914	0.2457	E.4	0.44
c. Interaction	2	$SS_{P_4} - (SS_{P4.A} + SS_{P4.S})$	3.2983	1.6492	E.4	2.93
D. Between Individual Means	18	$U_I - V_I$	37.5585	2.0866	E	2.74*
E. Between Individual Trends	(72)	$U_Y - U_I - V_Y + V_I$	(54.8890)	(0.7623)		—
1. Linear	18	$U_{P_1} - V_{P_1}$	7.7618	0.4312		—
2. Quadratic	18	$U_{P_2} - V_{P_2}$	16.4291	0.9127		—
3. Cubic	18	$U_{P_3} - V_{P_3}$	20.5520	1.1417		—
4. Quartic	18	$U_{P_4} - V_{P_4}$	10.1461	0.5636		—
F. Total	119	$U_Y - W_I$	165.9799	—		—

* Significant at 1% confidence level.
† Significant at 0.1% confidence level.

7, are a little difficult to justify, but probably the best error term for these would be the between-individual-trends mean square, line E of Table 7. These F tests have been made and are entered in the last column of Table 7.

The significant F's of Table 7 may be briefly interpreted. The significant over-all trend means simply that the over-all average scores vary from stage to stage during the test. Since only the linear component of the over-all trend is statistically significant, it may be concluded that the over-all trend is essentially linear. Higher order components of the over-all trend fail to attain significance. There are no significant differences between group means due to anxiety, shock, or the interaction of anxiety and shock. The general between-group-trends F is not significant. The linear components of the differences in group trends are, however, highly significant, and in particular, the anxious groups tend to have a less negative slope than the nonanxious groups. There are also significant differences between the different shock groups in the linear component. None of the higher order components of the differences between group trends was significant. (This was anticipated, but the tests were made to illustrate procedures.) There are significant individual differences in average performance as shown by the significant F between individual means, but this will usually be found in the case of reliable measures of performance.

The procedures outlined above can readily be applied to longer series of trials and can be extended to more complicated experimental designs with higher-order interactions, although in many instances interpretations will become very obscure. The number of subjects may vary from group to group, but if an orthogonal design is used the numbers of subjects in rows and columns must be proportional (**9**, pp. 281–**284**).

One other case which should be described briefly is that in which there is a single curve made up of average values from repeated measurement on a number of individuals. This would be the case if the Grant-Schiller data reported earlier had been obtained on a single group of subjects tested at all the different generalization test stimulus values. (Then each subject would have contributed a score for each generalization test stimulus.) In the case of repeated measurements the over-all trend can be analyzed as described above in Table 7. There will be no *between group differences* but *between individual means* and *between individual trends* measures can be obtained and separated into the orthogonal components. In this case the V column of Table 4 will not be computed, and the between-individual-means sum of squares will be $U_1 - W_1$, and the corresponding linear, quadratic, cubic, etc., terms will consist of the quantities from the W column subtracted from the corresponding quantities in the U column instead of the V's from the U's as in the present Table 7. The error terms will correspond to those used in the analysis in Table 7.

The above procedures do not constitute a universally applicable routine method of analyzing and comparing trends. No such method can exist, and no routine can substitute for an experimenter's insight and ingenuity. We have found the procedures extremely useful, however, in comparing curves with respect to slopes, curvatures, sharpness of inflections, etc., so that we can recommend them highly for testing specific hypotheses relating to trends.

REFERENCES

1. Alexander, H. W. A general test for trend. *Psychol. Bull.* 1946, **43**, 533–557.
2. Bartlett, M. S. The use of transformations. *Biometrics* 1947, **3**, 39–52.
3. Cochran, W. G., & Cox, Gertrude M. *Experimental designs.* New York: Wiley, 1950.
4. Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.
5. Edwards, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954. Pp. 330, 334.
6. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* (4th Ed.) New York: Hafner, 1953.
7. Grant, D. A., & Schiller, J. J. Generalization of the conditioned galvanic skin response to visual stimuli. *J. exp. Psychol.*, 1953, **46**, 309–313.
8. Pearson, E. S., & Hartley, H. O. *Biometrika tables for statisticians.* Cambridge: Cambridge Univer. Press, 1954.
9. Snedecor, G. W. *Statistical method.* (4th Ed.) Ames: Iowa State College Press, 1946.
10. Taylor, Janet A. A personality scale for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
11. Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.

Received April 4, 1955.

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 2, 1956

METHODS AND TERMINOLOGY IN STUDIES OF TIME ESTIMATION[1]

DALBIR BINDRA AND HÉLÈNE WAKSBERG
McGill University

In experiments on time estimation, *S* is required to make judgments about temporal durations. The duration that *S* is asked to judge is the *standard;* the estimate of the standard made by *S* is called the *judgment*. The studies in the area of time estimation are concerned primarily with the relative magnitudes of, and the general relations between, the standard and the corresponding judgment. Despite this relatively simple design of time estimation studies, they are quite confusing to read, and an attempt to make generalizations from the results of different studies gets one involved in many apparent or real contradictions. These troubles seem to arise from the fact that different investigators not only use different methods in their experiments, but also employ different sets of terminologies to describe their results. Thus, one investigator uses the method of reproduction in the study of a problem while another employs the method of verbal estimation for the same problem. Similarly, one researcher chooses to describe his results in terms of over- or underestimation of time, while another prefers to talk in terms of relative speed of internal and external clocks. This makes it difficult to compare studies, unless the exact relations between the different methods, and the different expressions used for describing results, have been clarified. Some investigators, for example Eson and Kafka (2) and Clausen (1), have pointed out the existence of this difficulty, but have dealt with it only in a limited way, within the framework of their particular problems. Even the standard textbooks of experimental psychology (for example, **5**) have not systematically examined the problem of the equivalence or lack of equivalence of the different methods and terms. This is the task of the present note.

METHODS

Three main methods are commonly used in time estimation experiments. In the method of *verbal estimation*, *E* delimits a given interval operatively (i.e., demonstrates the duration of the standard), and *S* is asked to estimate verbally its duration (the judgment) in terms of seconds or minutes. In the method of *production*, *S* is instructed to delimit operatively an interval (the judgment) of a given duration (the standard) stated verbally by *E*. In the method of *reproduction*, *E* operatively delimits an interval (the standard) and then asks *S* to reproduce operatively an interval (the judgment) of the same duration. (A variation of the method of reproduction is the method of *comparison*. In this method *E* presents two intervals consecutively and *S* is asked to judge their relative duration by saying which one is the longer. It resembles the method of reproduction in that both involve an operative presentation of the standard and a judgment that also refers to an opera-

[1] This study was supported by a research grant (A.P.12) from the National Research Council of Canada. It is based in part on an M.A. thesis, "Serial-position gradient in time estimation," submitted to McGill University, 1955.

tively delimited duration.) Thus we see that both the standard and the judgment may be defined in either of two different ways, by operative demonstration or by verbal statement. The term *elapsed time*, used by some authors, is used synonymously either with the standard or with the judgment, and does not require separate discussion here. Table 1 shows how the three methods differ with respect to the defining operations of the terms "standard," "judgment," and "elapsed time."

The three methods also differ with respect to another variable. These differences are also shown in Table 1. This variable is *objective* vs. *subjective* definition of the terms "standard," and "judgment." (Elapsed time is always defined in terms of objective time.) The standard is sometimes defined in terms of objective or clock time; at other times it is defined with reference to subjective or personal time. The judgment may also be defined in either one of these two ways. Thus, logically, we can combine standard and judgment in four different ways: (*a*) standard defined in terms of objective time and judgment defined in terms of subjective time; (*b*) standard defined in terms of subjective time and judgment defined in terms of objective time; (*c*) standard defined in terms of objective time and judgment defined in terms of objective time; and (*d*) standard defined in terms of subjective time and judgment defined in terms of subjective time. The last of these combinations is empirically meaningless. We have only three meaningful combinations and, correspondingly, three main methods of time estimation (Table 1). Thus we see that the different methods used in time estimation experiments represent the logical combinations of the two ways (in terms of objective vs. subjective time) of defining standard and judgment.

Description of Results

The results of time estimation experiments have been described in a number of different ways. "Over- or underestimation of the standard" (**5**), "over- or underestimation of elapsed time" (**2**, **3**), "the relative speed of the 'internal' and 'external' clocks," and "the relative magnitude of the subjective and objective temporal units" (**4**) are alternative ex-

TABLE 1

SIGNIFICANCE OF THE TERMS "STANDARD," "JUDGMENT," AND "ELAPSED TIME" IN VARIOUS METHODS OF TIME ESTIMATION

Terms	Methods		
	Verbal Estimation	Production	Reproduction
Standard	Interval delimited operatively by *E*; refers to objective (clock) time.	Interval stated verbally by *E*; refers to subjective time.	Interval delimited operatively by *E*; refers to objective time.
Judgment	Verbal estimate made by *S*; refers to subjective time.	Operative estimate made by *S*; refers to objective time.	Operative estimate made by *S*; refers to objective time.
Elapsed Time	Refers to the objective duration of the standard (operatively defined).	Refers to the objective duration of *S*'s judgment (operatively defined).	Refers to the objective duration of the standard, as well as of *S*'s judgment (operatively defined).

pressions that have been used by different investigators. Thus, a judgment larger than the standard obtained with the method of verbal estimation may be described as denoting overestimation of the standard, or overestimation of elapsed time, while a similar result obtained with the method of production may be described as denoting overestimation of the standard, or underestimation of elapsed time. Some investigators present their results only in terms of subjective and objective clocks or temporal units. The exact relations between all these expressions have not been adequately clarified. Table 2 attempts to show for the different methods the significance of these various expressions with respect to the relative magnitudes of the judgment and the standard.

Elapsed time refers to temporal durations as measured by standard clocks (objective or "external" clocks). "Internal clock" refers to a hypothethical mechanism in *S*, which is immediately and directly related to performance in a time estimation task. The rate of this clock may vary from individual to individual and in the same individual from time to

TABLE 2

SIGNIFICANCE OF THE RELATIONS BETWEEN THE JUDGMENT AND THE STANDARD IN VARIOUS METHODS OF TIME ESTIMATION. THE FOUR ENTRIES IN EACH BOX ARE EQUIVALENT STATEMENTS OF THE RELATION BETWEEN THE JUDGMENT AND THE STANDARD

Relative magnitude of judgment and standard	Methods		
	Verbal Estimation	Production	Reproduction
Judgments larger than the standard	1. Overestimation of the standard	1. Overestimation of the standard	1. Overestimation of the standard
	2. Overestimation of elapsed time	2. Underestimation of elapsed time	2. Not applicable
	3. Internal clock faster than external clock	3. Internal clock slower than external clock	3. Internal clock slower than external clock *during reproduction*
	4. Subjective temporal units smaller than objective temporal units	4. Subjective temporal units larger than objective temporal units	4. Subjective temporal units larger than objective temporal units *during reproduction*
Judgments smaller than the standard	1. Underestimation of the standard	1. Underestimation of the standard	1. Underestimation of the standard
	2. Underestimation of elapsed time	2. Overestimation of elapsed time	2. Not applicable
	3. Internal clock slower than external clock	3. Internal clock faster than external clock	3. Internal clock faster than external clock *during reproduction*
	4. Subjective temporal units larger than objective temporal units	4. Subjective temporal units smaller than objective temporal units	4. Subjective temporal units smaller than objective temporal units *during reproduction*

time. Subjective temporal units refer to subjective norms of the magnitude of the duration of seconds and minutes, and are presumably acquired through learning. These subjective units, like the internal clock, may be presumed to vary from individual to individual and in the same individual from occasion to occasion depending on internal and external conditions. It is clear that the concepts of subjective temporal unit and of internal clock are closely related.[2]

Three examples, one with each method, will help to clarify the relations shown in Table 2. Consider the situation in which *E*, using the method of estimation, presents a standard of 15 sec., and obtains from *S* a judgment of 20 sec. The judgment being larger than the standard, we say that *S* has overestimated the standard. Since in the method of estimation, the standard and elapsed time refer to the same thing (Table 1), that is, to the objective duration of the interval delimited by *E*, we can also say that *S* has overestimated elapsed time. The fact that *S* thought more seconds had elapsed than actually had, in the given duration of the standard, means that his subjective temporal units were smaller than objective temporal units. This is equivalent to saying that *S*'s internal clock ran faster than the objective clock.

Next, consider the situation in which *E*, using the method of production, obtains from *S* the same judgment of 20 sec., when the verbally stated standard is again 15 sec. Since the judgment is larger than the standard, we can say that *S* has overestimated the standard. Since, in the method of production, elapsed time refers to the duration of the interval delimited by *S* (Table 1), we can also say that he has underestimated elapsed time. The fact that *S* thought he had produced 15 sec. while he actually produced 20 sec. (since 15 of *S*'s subjective seconds elapsed in the course of 20 objective seconds) means also that *S*'s subjective temporal units were larger than objective temporal units. This is equivalent to saying that his internal clock ran at a slower rate than the objective clock.

Finally, consider the situation in which *E*, using the method of reproduction, delimits an interval of 15 sec. and requires *S* to reproduce an interval of the same duration. If *S*'s reproduction is 20 sec., his judgment is larger than the standard; he has overestimated the standard. But since, in the method of reproduction, elapsed time refers both to the duration delimited by *E* and the duration delimited by *S* (Table 1), we cannot make any meaningful statement regarding over- or underestimation of elapsed time. Further, in the method of reproduction, the relation between the magnitude of the judgment and the subjective temporal units or the rate of the internal clock is not as simple as in the other methods. In the other methods, over- or underestimation necessarily implies that the *S*'s subjective temporal units (and the rate of internal clock) is different from objective temporal units (and the rate of external clocks). But in the case of the method of reproduction, whether *S*'s internal clock runs faster or slower than the objective clock, he may still reproduce the duration of the standard quite accurately, for his subjective temporal units are not likely to change from the time that he is exposed to the standard to the time that he repro-

[2] The authors are grateful to Dr. Peter Milner for clarification of some of these relations.

duces it. However, if, as in our example, *S* does give a judgment larger than the standard, it can be said that *S*'s subjective temporal units were not only larger than objective units at the time of the reproduction (judgment), but were also larger than his subjective units at the time of the presentation of the standard. This is equivalent to saying that, during reproduction, his internal clock ran slower relative, not only to the external clock, but relative also to his own internal clock at the time of the presentation of the standard. In the method of reproduction, then, the relation between the magnitudes of judgment and standard cannot be as readily and simply described in terms of subjective temporal units or internal clock as in the case of the other methods.

Thus we see that a given difference between standard and judgment may signify quite different underlying events in the different methods. An identical difference may signify faster internal clock in the method of verbal estimation, slower internal clock in the method of production, and slowing down of the internal clock in the method of reproduction. Stating results in terms of the relative magnitudes of standard and judgment (over- or underestimation of the standard) is satisfactory only when all data are obtained with the same method. When, in the course of reviewing studies and making generalizations, it becomes necessary to compare results obtained with different methods and to theorize about the exact mechanisms underlying time estimation, it would seem desirable to restate the results in terms of the relative speeds of internal and external clocks or relative magnitudes of subjective and objective temporal units.

REFERENCES

1. Clausen, J. An evaluation of experimental methods of time estimation. *J. exp. Psychol.*, 1950, **40**, 756–761.
2. Eson, M. E., & Kafka, J. S. Diagnostic implications of a study in time perception. *J. gen. Psychol.*, 1952, **46**, 169–183.
3. Falk, J. L., & Bindra, D. Judgment of time as a function of serial position and stress. *J. exp. Psychol.*, 1954, **47**, 279–282.
4. Hoagland, H. The physiological control of judgments of duration: evidence for a chemical clock. *J. gen. Psychol.*, 1933, **9**, 267–287.
5. Woodrow, Herbert. Time perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1224–1236.

Received May 20, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

RELATIVE TASK DIFFICULTY AND TRANSFER OF TRAINING IN SKILLED PERFORMANCE

R. H. DAY
University of Sydney

A number of recent investigations have drawn attention to the part played by the relative difficulty of two or more tasks in the degree of transfer from one task to another. Results from these experiments have important implications for practical issues in training as well as for transfer theory. Since, however, there are certain inconsistencies in defining difficulty as well as discrepancies among the experimental results this area of investigation is in need of critical evaluation. The main purpose of this review, therefore, is to bring within the scope of one paper evidence from numerous experiments which have been directly or indirectly concerned with the effect of task difficulty on transfer in skilled performance and to subject this evidence to critical analysis. It should be mentioned in this connection that a number of the experiments dealt with here have not been primarily concerned with task difficulty in relation to transfer, but the evidence has arisen more or less incidentally during the course of the experiment. Also, this paper will be limited to a consideration of recent investigations within the field of human skill, although it is known that some earlier work (**6**) has drawn attention to the relationship existing between task difficulty and transfer of training.

Essentially, a skilled task in its simplest form possesses three basic features. These are: (*a*) a stimulus complex sometimes called the display, (*b*) devices by means of which elements in the display are brought under control by the responses of the operator, and (*c*) a linkage between these two. This broad analysis applies whether the task is a simple one such as writing with a pen upon paper, or a very much more complex task such as is involved in various kinds of tracking behavior. The difficulty of a task may vary as a function of changes in one or more of these features. The manner in which task difficulty has been varied within each of these task features will be treated first in this paper. This will be followed by a review of results obtained after which certain methodological and theoretical issues will be taken up.

For the most part the general procedure used in experiments of this nature involves either training a number of matched groups under different conditions of task difficulty followed by performance under a different condition, or the use of the AB, BA paradigm or an expansion of it in which each *S* undergoes each experimental condition. Initial task difficulty is usually assessed from the mean score for a number of trials, the score achieved on the final trial, or the mean number of trials required to reach a certain criterion. The method of estimating the extent of transfer also varies between studies.

Methods Used to Vary Task Difficulty

Stimulus Variations

Barch (**3**) using a following tracking task (Modified Two-Hand Coordination Test) varied difficulty by

changing target size. Using the same task Morin (**17**) varied difficulty in the same manner. A Rudder Control Test (Model CM120C) was employed by Gagné and Bilodeau (**9**). Variations in task difficulty were introduced by changing the width of the on-target scoring area. In an experiment by Szasfran and Welford (**23**) in which *S*s were required to throw loops of chain into a box the display was varied in three ways in order to obtain three levels of difficulty. Under the first conditions *S*s threw directly into the box, under the second they threw over a bar placed between *S* and target, and under the third condition a mirror was placed behind the box and a screen before it. Under the last condition it was necessary to use the mirror in aiming since the screen blocked direct viewing. In this experiment variations in display led necessarily to changes in method of throwing (responding) according as to whether *S* threw directly or over the bar. Alterations in target speed in a following tracking task led to variations in task difficulty in an experiment by Lincoln and Smith (**14**). Changes in target speed necessitated corresponding changes in speed of responding. In an experiment by Andreas *et al.* (**1**) task difficulty was varied by altering the number of moving elements in the display of a tracking task. This was accomplished by employing a following (SAM Two-Hand Pursuit Test) and a compensatory (SAM Two-Hand Coordination Test) task. Under the latter condition changes in display-control linkage were also introduced. A motor-discrimination task in which difficulty was varied by changing the nature of discrimination training during the initial task was used by Gagné *et al.* (**7**). Position and color discriminations only were used during the initial task and in the final phase both kinds of discrimination were involved.

Response Variations

Gibbs (**10**) used a hand-wheel controlled compensatory tracking task in which difficulty varied as a function of hand-wheel diameter and in which the smaller hand-wheel resulted in a more difficult task. Another method of varying task difficulty used by Gibbs (**10**) was that of changing the complexity of the path to be followed in a following tracking task. It is conceivable, of course, that a more complicated path provided greater perceptual difficulties as well as demanding more complex responses. A lever and a pressure control were also used to vary difficulty in another investigation by Gibbs (**11**) using a compensatory tracking task. The tracking task used by Baker *et al.* (**2**) was of the following variety in which variations in gear ratios between hand-wheel control and follower led to changes in the speed of movement of the follower relative to hand-wheel turning rate.

Variations in Control-Display Linkage

In the experiment by Barch (**3**) above, task difficulty was varied by means of a complete and a partial reversal of control-display relationship used in the standard form of the task. A change from a "natural" or "expected" to an "unnatural" or "unexpected" relationship resulted in an increase in task difficulty in an experiment by Gibbs (**10**). A following tracking task (Iowa Pursuit Apparatus) in which difficulty was varied in four ways using standard, reversed, and two partially reversed display-control linkages was employed by Barch and Lewis (**4**). Under the condition of partial display-control linkage reversal either the left or

right hand linkage was reversed. It is interesting to note that results from this experiment are contrary to those obtained by Gibbs (**10**). Lincoln (**13**) varied task difficulty by using direct, velocity, and aided controls.

Results of Experiments

At this point in the discussion it is as well to summarize briefly the general trend of the experimental results when task difficulty is varied in the numerous ways outlined in the previous section. The effects of relative task difficulty on transfer when difficulty is changed along a stimulus dimension are mainly negative when the *stimulus alone* is varied. In three experiments (**9**, **17**) greater initial task difficulty produced by changing the target dimensions did not give rise to a greater amount of transfer than when the initial task was relatively less difficult than the final task.[1] Target size, however, did exercise an effect in a fourth experiment (**3**). In the case of this last experiment it should be pointed out that linkage variables, the difficulty effects of which were not determined, could have affected the amount of transfer from initial to final task. Greater initial task tracking difficulty in terms of target speed (**14**) also failed to give greater transfer to an easier task than did the opposite order. Increasing task difficulty by changing from a following to a compensatory tracking task (**1**) did result in greater transfer when the difficult task was practiced initially, but only when the second task had an "unnatural" control-display arrangement. This effect was not observed when the final task control-display relationship was "natural" but still less difficult than the first task. In an aiming experiment (**23**) increasing the difficulty of the task by increasing the complexity of the display did result in greatest transfer in the difficult to easy direction. In this experiment, however, the variations in the perceptual aspects of the task led to changes in the mode of responding. This makes it difficult to attribute the differential transfer effects to stimulus variation alone. Such an argument applies, of course, to other experiments in which variations in the stimulus situation may have given rise to unmeasurable changes in the mode of response. It does appear, however, to be especially pertinent in the case of the aiming experiment. In a motor-discrimination task (**8**) transfer of training to a total task was greater when the more difficult of the two forms of discrimination was practiced first. It must be borne in mind, however, that in this experiment transfer of training from the components of a task to a total task comprising a combination of the two initial forms of the task was dealt with, rather than transfer from one task condition to another. The results from this investigation may not be altogether comparable with other experiments in which the stimulus situation was varied.

A greater degree of transfer from a difficult to a less difficult condition than from a less difficult to a difficult condition is met with more consistently when task difficulty is varied with respect to response variables. When the task was made more difficult by altering hand-wheel size and

[1] The results from an experiment by Green published after the completion of this paper in which target size was varied over a wide range are in agreement with these previous findings. The task used was a following tracking task. (Green, R. F., Transfer of skill on a following tracking task as a function of task difficulty (target size) *J. Psychol.*, 1955, **39**, 355–370.)

course complexity (**10**), method of control (**11**), and rate of responding (**2**), greatest transfer resulted when this form of the task was practiced initially.

Two experiments (**10**, **4**) have dealt with the effects on transfer of task difficulty varied with respect to the spatial relationship between control and display in tracking tasks. Whereas one of these investigations (**10**) has demonstrated that greater transfer occurred in the difficult to easy direction than in the opposite order, the other experiment (**4**) failed to note this effect. The fundamental differences in the design, extent of control-display relationship reversal, nature of task, and number of control devices must be considered as possible reasons for the inconsistency between these two sets of results. Variation in difficulty as a result of altering the nature of the control-display linkage in tracking (**13**) also failed to affect transfer of training differentially with respect to the degree of difficulty of initial and final tasks.

It is necessary to direct attention to the fact that the transfer phenomenon under consideration here has been investigated in connection with only a limited number of factors which are known to affect the difficulty level of a skilled task. Thus, with respect to response variables, such factors as hand-wheel inertia, friction, and aiding time-constants (**12**), relationship between display and control (**15**), control-crank radius (**22**), handedness (**21**), and planes of operation (**18**), have been systematically investigated in relation to the extent to which these factors affect ease of control. In the case of stimulus variables, target dimensions (**15**), visual magnification of target (**12**), structure of target surround (**20**), as well as others, have been examined in relation to task difficulty. A systematic examination of such variables from the point of view of differential transfer effects in terms of the order of the difficult and easy conditions would be of considerable practical value, as well as a contribution to an understanding of the transfer process in skilled performance.

Task Difficulty

The Isolation and Control of Variables

Ordinarily the relative difficulty of two or more tasks is defined in terms of the magnitude of mean scores achieved during whole or part of the training session or sessions, or, in terms of the magnitude of the score reached on a final trial. These scores are then subjected usually to statistical examination in order to establish the significance of differences between them. Should these differences prove to be significant, then that form of the task revealing greatest mean accuracy, least error, least time of performance, or some such, is said to be the least difficult. In transfer studies of the kind reviewed here, it is of importance not only to define operationally the relative difficulty of the tasks, but to establish as far as possible the source or cause of task difficulty differences. This is an essential step if a theory is to be constructed to deal with the effects on transfer of relative task difficulty.

A problem which has become plain from the preceding review is the difficulty of varying one task variable along a scale of difficulty without producing unmeasurable changes in other closely related variables. Even though the relative difficulty of the two tasks may have become obvious from an examination of performance

scores, the exact manner in which task difficulty arose, or to which task variable task difficulty can be attributed, has not always been clear. In short, in many of the experiments discussed the *locus* of task difficulty has not always been clearly specifiable. It should be remembered, however, that the stimulus or response conditions probably never remain the same from trial to trial, but, as Osgood (**19**) has been careful to show, stimulus and response may retain identical functions during learning. Even allowing for this fact many of the experiments on the effects of task difficulty on transfer have failed to control adequately the source of difficulty.

In the experiment by Gibbs (**10**) in which a following tracking or "steering" task was used not only course complexity, but also the complexity of the responses demanded by the course, varied under the two conditions. The more difficult of the two tasks presented the operator with a task which required responses of greater complexity than the easier task, as well as with a stimulus situation which may well have been more difficult perceptually. It is by no means easy in this case to state definitely the separate contributions of each of these factors to the difficulty of the task, or to the extent of transfer. A similar problem is met with in the investigation of Lincoln and Smith (**14**) where variations in target speed may well have given rise to greater perceptual difficulty, as well as greater difficulty in responding. In the experiment by Szasfran and Welford (**23**) this same problem arises again. The "bar" condition not only demanded an alteration in the response by requiring the subject to throw over it, but changed as well the stimulus situation. It is not possible, of course, to state whether or not the "bar" and "direct" conditions remained functionally identical while at the same time it is equally impossible to state definitely the locus of task difficulty. A similar argument applies when the greater difficulty of the "screen" condition in this investigation is considered. The two task conditions employed by Andreas *et al.* (**1**) varied not only with respect to the number of display elements, but also in relation to course and response complexity and control-display relationships. In this case it is not possible to state the individual contributions to task difficulty nor their possible interactions, all of which could have affected the level of difficulty.

A further kind of problem in the control of variables has arisen when changes along one stimulus or response dimension have led to unmeasurable variations along a closely related dimension. For example, Baker *et al.* (**2**) altered the level of task difficulty by varying the rate of hand-wheel turning necessary to move the target-follower through a certain distance. These authors have pointed out that changes in turning rate altered also the extent of movement as well as the required force of movement. Variations in task difficulty could be due to any one of these factors or to an interaction between two or all of these.

The principal problem, then, in many of the studies reviewed is to state the locus of variations in task difficulty. Without doing this transfer of training of skill as a function of the relative difficulty of the tasks is not easy to deal with theoretically, since the actual experimental evidence remains obscure. The problem of isolation and control of "difficulty" variables in skilled tasks does not lend itself easily to solution, since

stimulus and response factors are so intimately related, and therefore exceptionally difficult to isolate under experimental conditions.

Task Difficulty and Performance Standards

In a number of skilled tasks the level of performance required is frequently *implied* rather than demanded by the task itself. This point can be made clearer by illustration. If in a following tracking task the target is 1 in. in diameter, the target follower ⅛ in. in diameter, and the task of the operator is to superimpose the latter upon the former, a greater margin of error is permitted than if both these display elements are ⅛ in. in diameter. The extent of permissible error is implied by the relative sizes of target and follower. In this example, whereas holding the follower superimposed upon the small (⅛ in.) target may not be necessarily a more difficult task than keeping it on a larger target, the structure of the stimulus situation in the latter case is such that there is a wider margin within which the off-target extent does not count as an error. This means that the two conditions mentioned differ with respect to operationally defined "difficulty" insofar as one condition demands higher performance standards rather than a higher level of skill. The more "difficult" task under these circumstances, then, is the one which requires a higher standard of performance by setting narrower error-tolerance limits. The two task conditions do not necessarily differ, or differ only to a small degree, in the actual level of skill which they demand for adequate performance. The differences, sometimes large, between performance curves for two task conditions differing in this manner may be due primarily to the fact that the operator is directing to the task very little, or a great deal of effort, since the task implies by its target and follower dimensions a certain standard of performance. The difficulty differences between the task conditions is more apparent than real.

Szasfran and Welford (**23**) have suggested that one possible explanation of the phenomenon of greater transfer from a difficult to an easy task may be found in the higher standards of performance established during the difficult initial task, and carried over to the easier final task. Transfer in this case would be expected to be greater than in shifting from an easy initial task to a more difficult final task. The available experimental evidence does little to support this hypothesis. In the experiments reported (**9**, **17**) dealing with the relative dimensions of target and follower in following tracking tasks in which the task difficulty was varied by changing error tolerance limits, transfer was positive and about equal in going from difficult to easy and from easy to difficult conditions. In another experiment (**3**) greater transfer was found in shifting from difficult to easy, but it has been pointed out that control-display relationship factors were confounded with stimulus factors in this study and the experimental design considered only the easy to difficult direction.

The concept of error-tolerance implicit to the task as a determinant of performance is not a new one. Mace (**16**) has put forward an hypothesis of "implicit standards" which is summarized in the statement " . . . subjects aiming at targets defined to themselves a 'good,' 'fair' or 'poor' shot not in terms of its *absolute* distance from the bull's-eye, but in a way which was *relative to the form of*

target employed" (**16**, p. 103). A similar view has been expressed by Helson (**12**) in his hypothesis of par or tolerance.

The present indications are that when the difficulty of a task varies only in terms of the implied extent of error tolerated, neither condition calling for a higher level of skill, the phenomenon of greater transfer from the difficult to the less difficult condition than from the less to the more difficult condition does not occur. So far this has only been observed in relation to the relative dimensions of target and follower in a tracking task. Much needs to be done with respect to both a variety of skilled tasks, and the various features of skilled tasks, before definite conclusions can be drawn or hypotheses clearly formulated.

The **U** *Hypothesis and Transfer*

Bartlett (**5**) and Helson (**12**) have each outlined hypotheses which state in effect that performance will remain essentially the same over a certain range of variation in the physical characteristics of the task. Outside this range performance tends to undergo considerable changes. Thus Bartlett states: "The fundamental features of performance will remain stable over a certain range of its conditions. Outside this range they will change often in a dramatic and radical manner" (**5**, p. 444). Helson's statement is much the same:

> Human performance tends to be optimal as judged by accuracy, efficiency, and comfort, over a more or less broad band of values for a given stimulus variable outside of which it becomes noticeably poorer. When performance is plotted in terms of error or the reciprocal of accuracy, the resultant curve is roughly **U**-shaped (**12**, p. 493).

Helson has demonstrated such a curve for aiding time-constants, hand-wheel turning speed, and hand-wheel size and inertia.

Employing this notion of the optimal band of performance Gibbs (**10**) has suggested that:

> . . . transfer between two tasks, one of which lies within the tolerance limits, and one of which lies outside this optimal zone can be (a) positive, (b) large, (c) unequal and unaffected by the order of task presentation; this implies greater transfer when the first task lies outside the optimal zone, and the second lies within it, than when the first task comes within and the second outside.

This hypothesis lends itself readily to experiment, and since a considerable amount of data is now available concerning the optimal band of performance for a number of task variables, experiments designed to measure transfer from within to without and from without to within the optimal zone would be of great value.

A single **U** curve deals only with a single variable thus emphasizing again the need for careful isolation and control of task variables in transfer experiments of this kind. Since much of the evidence indicates that the relative difficulty of initial and final task conditions is a transfer determinant of considerable importance it is essential now to study the many factors varying in difficulty and depicted by the **U** shaped curve. This approach would doubtlessly provide data of fundamental importance to practical and theoretical considerations in the transfer of skill.

In conclusion, one further inadequacy in the experimental design of many of the experiments so far mentioned needs to be pointed out. Many of the experiments have been designed so that only the difficult-to-easy, and easy-to-difficult task conditions have been taken into consideration. It is again important, from both practical and theoretical viewpoints, that the degree of transfer from one condition to the same con-

dition should be observed as well. In the experiment by Baker *et al.* (2) it was found that the greatest degree of transfer occurred when there was no difference in difficulty between the initial and final task. This means, of course, that greatest facilitation took place when the initial and final tasks were the same task. This is normal learning rather than transfer of skill. Since a large number of the investigations outlined in this paper failed to include the experimental conditions of easy to easy and difficult to difficult, it is not easy to generalize concerning the difficulty conditions of initial and final task for maximum transfer.

Summary

This article has presented a summary of a number of investigations concerned with the effect on transfer of training of the relative difficulty of initial and final tasks. The results from a number of recent studies are regarded as presenting an important problem for practical consideration and theoretical interpretation. The principal findings of these experiments have been briefly summarized. The concept of task difficulty has been discussed in relation to the isolation and control of task variables, subjective performance standards, and the **U** hypothesis.

REFERENCES

1. Andreas, B. G., Green, R. F., & Spragg, S. D. S. Transfer effects between performance on a following tracking task (Modified SAM Two-Hand Coördination Test) and a compensatory tracking task (Modified SAM Two-Hand Pursuit Test). *J. Psychol.*, 1954, **37**, 173–183.
2. Baker, Katherine E., Wylie, R. C., & Gagné, R. M. Transfer of training to a motor skill as a function of variation in rate of response. *J. exp. Psychol.*, 1950, **40**, 721–732.
3. Barch, A. M. The effect of difficulty of task on proactive facilitation and interference. *J. exp. Psychol.*, 1953, **46**, 37–42.
4. Barch, A. M., & Lewis, D. The effect of task difficulty and amount of practice on proactive transfer. *J. exp. Psychol.*, 1954, **48**, 134–142.
5. Bartlett, F. C. Some problems of "display" and "control." In A. Michotte (Ed.), *Miscellanea psychologica.* Louvain: Librairie Philosophique, 1947. Pp. 440–452.
6. Cook, T. W. Amount of material and difficulty of problem solving. *J. exp. Psychol.*, 1937, **20**, 288–296.
7. Gagné, R. M., Baker, Katherine E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
8. Gagné, R. M., Baker, Katherine E., & Foster, H. Transfer of discrimination training to a motor task. *J. exp. Psychol.*, 1950, **40**, 314–328.
9. Gagné, R. M., & Bilodeau, E. A. The effects of target size variation on skill acquisition. *USAF, Personnel Train. Res. Cent., Res. Bull.*, 1954, No. 54-55.
10. Gibbs, C. B. Transfer of training and skill assumptions in tracking. *Quart. J. exp. Psychol.*, 1951, **3**, 99–110.
11. Gibbs, C. B. The continuous regulation of skilled response by kinaesthetic feed back. *Med. Res. Coun., Appl. Psychol. Unit. Rep.* 1953, No. A.P.U. 190/53.
12. Helson, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **62**, 473–479.
13. Lincoln, R. S. Visual tracking III. The instrumental dimension of motion in relation to tracking accuracy. *J. appl. Psychol.*, 1953, **37**, 489–493.
14. Lincoln, R. S., & Smith, K. U. Transfer of training in tracking performance at different target speeds. *J. appl. Psychol.*, 1951, **35**, 358–362.
15. Lincoln, R. S., & Smith, K. U. Systematic analysis of factors determining accuracy in visual tracking. *Science*, 1952, **116**, 183–187.
16. Mace, C. A. The influence of indirect incentives upon the accuracy of skilled movements. *Brit. J. Psychol.*, 1931, **22**, 101–114.
17. Morin, R. E. Transfer of training between motor tasks varying in precision of movement required to score. *Amer. Psychologist*, 1951, **6**, 390. (Abstract)

18. NORRIS, EUGENIA B., & SPRAGG, S. D. S. Performance on a following tracking task (Modified SAM Two-Hand Coordination Test) as a function of the planes of operation of the controls. *J. Psychol.*, 1953, **35**, 107–117.

19. OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychol. Rev.*, 1949, **56**, 132–143.

20. SENDERS, J. W. The influence of surround on tracking performance. Pt. 1. Tracking on combined pursuit and compensatory one-dimensional tasks with and without structural surround. *USAF, WADC Tech. Rep.* 1953 No. 52-229.

21. SIMON, J. R., DE CROW, T. W., LINCOLN, R. S., & SMITH, K. U. Effects of handedness on tracking accuracy. *Percept. Mot. Skills. Res. Exch.*, 1952, **4**, 53–57.

22. SWARTZ, P., NORRIS, EUGENIA B., & SPRAGG, S. D. S. Performance on a following tracking task (Modified SAM Two-Hand Coordination Test) as a function of radius of control cranks. *J. Psychol.*, 1954, **37**, 163–171.

23. SZASFRAN, J., & WELFORD, A. T. On the relation between transfer and difficulty of initial task. *Quart. J. exp. Psychol.*, 1950, **2**, 88–94.

Received March 21, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

EYSENCK'S TENDER-MINDEDNESS DIMENSION: A CRITIQUE

MILTON ROKEACH and CHARLES HANLEY
Michigan State University

In a series of articles and books (**1**, **2**, **3**, **4**, **5**, **6**), Eysenck has tried to demonstrate that individual differences in social attitudes are reducible to "two primary social attitudes." The first of these he calls "radicalism-conservatism" and the second, "tough-mindedness–tender-mindedness."[1] Our primary concern in this paper is with the empirical and theoretical bases of the latter of these dimensions.

The origin of the two factors has been described by Eysenck:

> This scheme was derived by the writer from a factorial analysis of the responses to a 40-item attitude inventory made by 750 middle-class English subjects. . . . These were drawn equally from voters for the three major British political parties (conservative, liberal, and socialist), and the 250 subjects representing each party were equated for age, sex, and education. It was shown that items having high factor saturations on radicalism-conservatism (R) also distinguished at a high level of significance between voters for the conservative and radical parties respectively, while items having low saturations on R failed to distinguish between the two parties. . . .
>
> The tender-mindedness factor (T) was found to be quite uncorrelated with R, and to give no discrimination between the political parties. On analyzing responses of additonal samples of communist and fascist party members, however, it was found that both these groups were differentiated with considerable accuracy from members of the three democratic parties by their low T scores; both communists and fascists were thus found to be tough-minded in comparison with conservatives, liberals, and socialists (**3**, p. 563).

Eysenck describes the bipolar tender-mindedness–tough-mindedness factor primarily in terms of the items which measure it (**1**). The reader may refer to the left-hand column of Table 1 for the 14 T items and the direction in which they are to be scored. In attempting to find an underlying concept corresponding to this factor, Eysenck uses William James's dichotomy of tender-mindedness vs. tough-mindedness. Eysenck notes "one set of opinions appears to be dominated by ethical, moralistic, super-ego, altruistic values; the other by realistic, worldly, egotistic values" (**1**, p. 61).

An examination of Eysenck's writings reveals that he explains, or reinterprets, a great body of research on the organization of social attitudes and their relation to personality in terms of the R and T factors. Thus, a considerable portion of his recent book, *The Psychology of Politics* (**5**), is devoted to just this undertaking.

In view of the importance which Eysenck attaches to the factor he labels tender-mindedness, and the far-reaching conclusions he draws regarding differences on this dimension among adherents to various political positions ranging from the Right to the Left, we felt that a careful scrutiny of his data was in order. On the basis of our analysis, we have arrived at the opinion that the evidence Eysenck himself presents to support these conclusions contains grave contradictions and errors in computation, and that the concept itself, at least as employed by Eysenck, contributes little or nothing to the scientific understanding of the social phenomena to which he addresses himself.

[1] It should be understood that whenever this term is used, we follow strictly Eysenck's own usage.

TABLE 1

FREQUENCY OF RESPONSES BY CONSERVATIVES, LIBERALS, SOCIALISTS, AND COMMUNISTS TO THE ITEMS ON THE TENDER-MINDEDNESS SCALE

Item and direction of scoring	Class	Proportion of "yes" answers			
		Cons.	Lib.	Soc.	Com.
Items on which communists are most tender-minded:					
1. Coloured people are innately inferior to white people. (T −)	Middle	.42	.27	.19	.00
	Working	.30	.33	.16	.06
3. War is inherent in human nature. (T −)	Middle	.67	.57	.34	.02
	Working	.70	.67	.60	.02
5. Persons with serious hereditary defects and diseases should be compulsorily sterilized. (T −)	Middle	.69	.59	.63	.46
	Working	.96	.83	.89	.71
8. In the interests of peace, we should give up part of our national sovereignty. (T +)	Middle	.32	.60	.76	.74
	Working	.37	.38	.50	.65
10. It is wrong that men should be permitted greater sexual freedom than women by society. (T +)	Middle	.66	.71	.80	.93
	Working	.74	.78	.76	.91
13. Conscientious objectors are traitors to their country, and should be treated accordingly. (T −)	Middle	.28	.16	.09	.02
	Working	.67	.22	.27	.06
36. The death penalty is barbaric, and should be abolished. (T +)	Middle	.30	.42	.64	.96
	Working	.19	.11	.20	.83
39. The Japanese are by nature a cruel people. (T −)	Middle	.58	.37	.19	.00
	Working	.74	.44	.27	.06
Items on which communists are most tough-minded:					
9. Sunday-observance is old-fashioned, and should cease to govern our behaviour. (T −)	Middle	.36	.44	.68	.92
	Working	.59	.33	.69	.95
15. The laws against abortion should be abolished. (T −)	Middle	.28	.40	.53	.90
	Working	.33	.11	.51	.65
16. Only by going back to religion can civilization hope to survive. (T +)	Middle	.65	.56	.36	.00
	Working	.74	.61	.27	.05
23. Divorce laws should be altered to make divorce easier. (T −)	Middle	.33	.42	.61	.96
	Working	.37	.22	.53	.91
28. It is right and proper that religious education in schools should be compulsory. (T +)	Middle	.66	.55	.32	.00
	Working	.70	.78	.13	.05
29. Men and women have the right to find out whether they are sexually suited before marriage (e.g., by companionate marriage). (T −)	Middle	.35	.40	.62	.98
	Working	.37	.22	.36	.77

The basis for our opinion rests primarily on an examination of the data which Eysenck presents in two papers which form the foundation of his work with the T factor. In the earlier of these publications (**1**), Eysenck reports the factorial analysis alluded to earlier, and in the second (**2**), he presents data regarding the performance of middle- and working-class members of the conservative, liberal, socialist, and communist parties. The mean values of T for these groups, as given by Eysenck (**2**), are shown in Table 2 under the heading "Original mean."[2]

TABLE 2

MEAN TENDER-MINDEDNESS SCORES OF MEMBERS OF VARIOUS BRITISH POLITICAL PARTIES AND SOCIO-ECONOMIC CLASSES

Party	Class	*N*	Original mean	Recomputed mean
Conservative	Middle	250	7.6	7.6
	Working	65	6.3	6.7
Liberal	Middle	250	7.9	8.2
	Working	27	7.4	8.3
Socialist	Middle	250	8.0	8.0
	Working	45	6.2	6.6
Communist	Middle	50	6.8	7.4
	Working	96	6.0	7.3

There are 14 items on the T scale; the subject responds to each item with a ++, +, 0, −, or −−, according to the degree of his agreement or disagreement with the content of the item. If the subject responds to all items in the tender-minded direction (e.g., −−, or −, to Item 1 in our Table 1), he obtains a score of 14. If, on the other hand, he answers all 14 items either in the tough-minded direction, or with "0," he gets a score of zero on the scale. From Table 2 it is seen that, according to Eysenck, the middle- and working-class communists are the least tender-minded of the four political groups. He indicates further than this difference is statistically significant.

It is reasonable to suppose, in view of these means, that the communists would tend to show up as least tender-minded on each of the 14 items measuring the T dimension. Fortunately, Eysenck has presented (**2**, p. 203) the percentage of agreement with each of the 14 T items, broken down according to socioeconomic and political groupings. On inspecting these data, we found that the communists were the most tough-minded of the groups on six of the 14 items, *but were the most tender-minded of all on the remaining eight items.* These findings, furthermore, obtained for both the middle- and working-class samples. In our Table 1, we have reproduced Eysenck's data on this point, rearranging it into two sections: (*a*) items on which communists were the most tender-minded, and (*b*) items on which communists were the most tough-minded.

It should come as no surprise to find that British communists are most tough-minded on six of the 14 T items, if by tough-minded one means that they are opposed to Sunday observance, favor the abolition of laws against abortion, are anti-religion, believe that divorce laws should be liberalized, are against compulsory religious education in the schools, and favor companionate marriage. Nor does it surprise us to find these same communists are most tender-minded on the remaining eight items, if by tender-minded one means that they reject the idea that colored people are inferior, disagree with the notion that war is inherent in human

[2] Means for seven fascists were also presented, unaccompanied by any breakdown of responses to individual items. Hence, we omit consideration of this small sample.

nature, oppose compulsory sterilization, are willing to give up national sovereignty, reject the double standard, oppose the idea that conscientious objectors are traitors, favor the abolition of the death penalty, and reject the proposal that the Japanese are inherently cruel. The conclusions we draw from these results, however, are considerably at variance from those Eysenck arrives at on the basis of differences in means on T (cf. "Original mean" in Table 2). Furthermore, we note Eysenck in coming to his conclusions makes no reference whatever to these data on the individual items. This oversight is in sharp contrast to the fact that he does not hesitate to point to such group differences in response to individual items as supporting the validity of the radicalism-conservatism factor (1, p. 60–61; see also our quotation from Eysenck at the beginning of this paper).

It is seen then, that while middle- and working-class communists are the most tender-minded of all the groups on eight of the 14 items, Eysenck's mean scores on T show them as significantly more tough-minded than conservatives, liberals, and socialists. It was difficult for us to reconcile such findings; therefore, we proceeded to recompute the means for the various samples. This is easily done using the information contained in Table 1 on the percentage of agreement with each item. The simplest way of doing this is to add the percentage values of tender-minded responses for all 14 items for a given group. It is necessary, of course, to keep in mind the direction of the scoring. For items on which disagreement indicates tender-mindedness, the percentage frequencies must be subtracted from 100 per cent in order to get the correct value to use in recomputing the means. Thus, the mean for the middle-class, conservative group is recomputed by summing the following tender-mindedness proportions obtained from the data in Table 1: (1, T−) .58; (3, T−) .33; (5, T−) .31; (8, T+) .32; (10, T+) .66; (13, T−) .72; (36, T+) .30; (39, T−) .42; (9, T−) .64; (15, T−) .72; (16, T+) .65; (23, T−) .67; (28, T+) .66; and (29, T−) .65. Their sum rounds to 7.6, the mean of the middle-class, conservative sample. The same value may be obtained with a bit more labor by the following procedure: (*a*) Find the number of tender-minded responses to each item by multiplying each of the above proportions by 250, the number of *Ss* in the conservative group; (*b*) sum these values; (*c*) divide this sum by 250.

The values obtained from our recomputation are shown in the "Recomputed mean" column of Table 2. A comparison of our means with Eysenck's reveals the following:

1. In only two out of eight comparisons are the means identical.

2. Both middle- and working-class communist means shift markedly toward the tender-minded direction (i.e., toward the values found for the other samples). In the case of the middle-class communists, the shift is from 6.8 to 7.4; with the working-class communists the change is even more dramatic, from 6.0 to 7.3.

3. The order of the middle-class groups in relation to tender-mindedness is unchanged, but the difference between communists and conservatives is indeed slight, 7.4 vs. 7.6.

4. With the working-class groups, the recomputed means indicate that the communists are more tender-minded than either conservatives or socialists, a drastic change from Eysenck's original finding.

It is conceivable that the discrep-

ancies appearing in Table 2 arise because our method of scoring diverges in one respect from Eysenck's. He *always* scores "0" responses as tough-minded, while we, because Table 1 gives only frequency of agreement, are forced to treat "0" answers to certain items (e.g., Item 1 and all others scored in the negative direction) as tender-minded. We do not believe that this difference in scoring technique accounts for the discrepancies. Assume first that the frequency of "0" responses is the same in all political groups. Then, the recomputed means should (*a*) be uniformly higher than those Eysenck reports, and the magnitude of increase should be approximately the same for all samples; and (*b*) the order of the groups on the T dimension should remain the same. Neither of the preceding occurs. With respect to *a*, in the middle-class groups the recomputed means for the conservatives and socialists show no change at all; further, the magnitude of increase in means is greatest in both middle- and working-class groups for the communists. With respect to *b* above, the order of the groups is strikingly altered in the working-class samples. Thus, it is clear that the discrepancies cannot be explained if we assume equal frequencies of "0" responses.

Another possibility comes to mind. Perhaps the communists show the greatest increases in recomputed means because they give relatively more "0" responses than do the other groups. This could account for the discrepancies. There is only indirect evidence bearing on this point, and it leads us to doubt that such is the case. Eysenck states that extreme responses to the items are *more* common in his communist samples. Thus, the communists have "a greater tendency to believe *strongly* in the correctness of the attitude held" (**5**, p. 140). We must conclude, therefore, that the discrepancies between our means and Eysenck's cannot be explained away by appealing to the manner in which we have had to handle "0" responses.

In view of the foregoing analysis, Eysenck's continued contention (**2**, **3**, **4**, **5**) that communists are more tough-minded than conservatives, liberals, and socialists, is not supported by his published data.

The Nature of the Tender-Mindedness Items

We do not consider it constructive to terminate our analysis at this point. What conclusions *can* be validly drawn from these data?

When one plots the factor saturations given by Eysenck (**1**) for the 14 items, it is immediately apparent that a rotation of approximately 45 degrees produces a striking increase in the number of items with near-zero saturations on one of the two rotated factors. Inspection of the items with high positive or high negative saturations on these rotated factors indicates clearly that two kinds of content are involved which are strikingly similar to Ferguson's factors of "religionism" and "humanitarianism" (**7**) and Kirkpatrick's dimensions of "religiosity" and "humanitarianism" (**8**).[3]

The positive pole of the "religionism" factor is indicated by Items 16 (going back to religion) and 28 (compulsory religious education), the negative pole by Items 9 (Sunday observance old-fashioned), 15 (abolish laws against abortion), 23 (liberalize divorce laws), and 29 (companionate

[3] The remainder of the items on Eysenck's questionnaire also were plotted in this way. The interpretation of the two rotated factors in no way needs to be altered to account for the content of these additional items.

marriage). It develops that the six T items on which communists are the most "tough-minded" are just these six "religionism" items. The communists take a clearly antireligious position.

The positive pole of the "humanitarianism" factor is established by Items 8 (give up national sovereignty) and 36 (abolish death penalty), the negative pole by Items 1 (colored people inferior), 3 (war inherent in human nature), 13 (conscientious objectors are traitors), and 39 (Japanese cruel by nature). One additional item, 10 (oppose double standard), has a moderate, positive "humanitarianism" saturation, while Item 5 (compulsory sterilization) receives a moderate, negative saturation. It turns out that these are the eight items on which the communists score as the most tender-minded of the various samples.

These rotated factors make it far more understandable why the communists score the most tender-minded on eight of the 14 T items. All eight are saturated with "humanitarianism." It is generally known that communist ideology supports the attitudes expressed by the positive pole and opposes the attitudes expressed by the negative pole of this factor. Similarly, it is easily understandable why the communists score as the most tough-minded on the remaining six T items. All six are saturated with "religionism." Again, it is consistent with communist ideology to agree with statements unfavorable to religion and to disagree with statements based on religious doctrine.

When scores on the rotated factors are computed for the various groups, using the percentages given in Table 1, the communists are highest on the "humanitarianism" factor, followed in order by socialists, liberals, and lastly, by conservatives. Conversely, on the "religionism" factor, the conservatives score highest among middle-class samples, followed in descending order by liberals, socialists, and finally, communists. The order in the working-class samples is similar, except that liberals score higher on "religionism" than do conservatives. Furthermore, the order of the frequency of responses of the four middle-class political groups to all but two items corresponds exactly to the positions of these parties on the Right-Left political axis (cf. Table 1). A strong tendency of a similar sort is apparent in the working-class groups, but it must be remembered that sampling errors are larger here because of the smaller number of subjects involved. Thus, most, if not all, of the items on the tender-mindedness scale are clearly related to political affiliation.

The preceding reanalysis of Eysenck's data in terms of "religionism" and "humanitarianism" meshes neatly with the earlier research by Ferguson (**7**) as well as that by Kirkpatrick (**8**). Eysenck did not mention Ferguson's 1941 paper in his original publications on the R and T factors (**1**, **2**), but in *The Structure of Human Personality* (**4**) and in *The Psychology of Politics* (**5**) he raises the question as to whether "radicalism-conservatism" and "tough-tender-mindedness" are superior to Ferguson's "religionism" and "humanitarianism" in accounting for the data. Eysenck decides in favor of his own dimensions. It is instructive to follow his reasoning as it applies to the problem. First, he argues that "radicalism" and "tender-mindedness" are to be preferred on the grounds of

"semantic convenience." "*More convincing,*" Eysenck continues, "*would be experimental evidence showing that Tough-mindedness had correlates in other fields, such as, for instance, in the field of personality, which neither Religionism nor Humanitarianism possessed. A proof of this type will be attempted in a later chapter, and the reader is asked to suspend his judgment until then* . . . " (**5**, p. 147).[4]

Such evidence would indeed be instructive. We made a careful search of the remainder of *The Psychology of Politics* for this promised experimental test. Our search was in vain. The issue is never again raised in the book. This sort of treatment, of course, can hardly be considered an adequate resolution of the problem and serves only to reinforce the conclusions we have reached.

Summary

This paper is an evaluation of Eysenck's research on the factor he calls tough-mindedness–tender-mindedness. Using a 14-item T scale dealing with such diverse religious issues as Sunday observance, abortion, divorce, companionate marriage, etc., and with such other social issues as race differences, the cruelty of the Japanese, compulsory sterilization, the double standard, and conscientious objectors, Eysenck reports mean scores which indicate that middle- and working-class British communists are more tough-minded than middle- and working-class British conservatives, liberals, and socialists.

Our analysis of Eysenck's published data clearly contradicts his findings and conclusions. The major points considered were:

1. Contrary to Eysenck's contention that communists are the most tough-minded, we find that on 8 of 14 items both middle- and working-class communists were the most tender-minded of all the political groups. Eysenck overlooks these data in arriving at his conclusions.

2. Our recomputation of the mean tender-mindedness scores, from data presented by Eysenck, strongly suggests serious errors in the original calculation of his means.

3. The corrected means and our analysis of Eysenck's data on frequency of agreement to individual items necessitate a serious modification of his conclusions regarding differences in tender-mindedness among the various political groups.

4. The 8 (out of 14) tender-mindedness items on which the communists were the most tender-minded of all parties turn out to involve content which corresponds closely to Ferguson's (**7**) and Kirkpatrick's (**8**) dimensions of humanitarianism.

5. The remaining six items on which the communists were the most tough-minded all pertain to a religious dimension, also found by Ferguson and by Kirkpatrick.

6. Consistent with common knowledge, Eysenck's conservatives score the highest on the religious items, followed in order by liberals, socialists, and communists. Communists score the highest on the humanitarianism items, followed in order by socialists, liberals, and conservatives.

Our analysis leads us to the conclusion that tough-mindedness–tender-mindedness, as conceived and measured by Eysenck, has no basis in fact. It is based on miscalculations and a disregard for a significant portion of his data. It conceals rather than reveals the attitudinal differences existing among political groups.

[4] Italics ours.

REFERENCES

1. Eysenck, H. J. Primary social attitudes: I. The organization and measurement of social attitudes. *Int. J. Opin. Attitude Res.*, 1947, **1**, 49–84.
2. Eysenck, H. J. Primary social attitudes as related to social class and political party. *Brit. J. Sociol.*, 1951, **2**, 198–209.
3. Eysenck, H. J. Primary social attitudes: A comparison of attitude patterns in England, Germany, and Sweden. *J. abnorm. soc. Psychol.*, 1953, **48**, 563–568.
4. Eysenck, H. J. *The structure of human personality.* London: Methuen, 1953.
5. Eysenck, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
6. Eysenck, H. J., & Crown, S. National stereotypes: an experimental and methodological study. *Int. J. Opin. Attitude Res.*, 1948, **2**, 26–37.
7. Ferguson, L. W. The stability of the primary social attitudes. I. Religionism and humanitarianism. *J. Psychol.*, 1941, **12**, 283–288.
8. Kirkpatrick, C. Religion and humanitarianism: a study of institutional implications. *Psychol. Monogr.*, 1949, **63**, No. 9 (Whole No. 304).

Received February 24, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

THE PSYCHOLOGY OF POLITICS: A REPLY

H. J. EYSENCK
Institute of Psychiatry
Maudsley Hospital, University of London

In the *Psychology of Politics* (**11**) and in a number of earlier articles and papers the writer has tried to do three things. In the first place, he has tried to construct a dimensional framework to deal with the interrelations obtaining between a wide variety of different social attitudes. The results of several experiments and analyses, carried out in different countries and on different samples, led to the hypothesis that these relationships could be described with considerable accuracy in terms of two orthogonal (independent) factors, labeled radicalism-conservatism (R factor) and tough-mindedness–tender-mindedness (T factor). No attempt was made, as Rokeach and Hanley (**16**) claim, "to demonstrate that individual differences in social attitudes are reducible to 'two primary social attitudes'"; such a reduction would fail to take account of the specific part of the variance, which is considerable, and could not be effected by the use of the factorial method, on which our conclusions were based.

In the second place, an attempt was made to follow up a hypothesis formulated quite early in the history of this research (**5**), to the effect that the T dimension was correlated with certain personality variables, while no such correlation was postulated for the R dimension. The specific hypothesis tested was that introverted people would tend to be tender-minded, while extraverted people would tend to be tough-minded. In this connection, the hypothetical constructs "introversion" and "extraversion" are used in terms of the operational definition given them in *Dimensions of Personality* (**7**), *The Scientific Study of Personality* (**9**), and *The Structure of Human Personality* (**10**).

In the third place, an attempt was made to link up both the attitude dimensions and the personality studies with the main body of modern psychology by showing that the results found in our experiments could be deduced from certain postulates of learning theory, and that in this way the particular structuring of variables observed could be explained by reference to a larger body of well-known facts. The claim is made in the *Psychology of Politics* (**11**) that these three aims have been accomplished to a reasonable approximation. In view of the fact that if this claim could be substantiated the work reported would be of some interest to social psychologists concerned with the integration of their field of study with that of general and experimental psychology, well-considered criticism showing possible weaknesses in the chain of proof is welcomed by the writer, as this would make possible the design of more convincing experiments, or lead to a more accurate restatement of the theory. It is to be regretted that the critique by Rokeach and Hanley (**16**) does not seem to be related closely enough to the facts of the case to be useful from this point of view.

Their first point of criticism appears to be that in one paper (**8**) the writer concluded that the communist groups tested had low scores on ten-

der-mindedness; this they claim to be an error based on miscalculation. Computational errors do, of course, occur even when considerable care is taken. The writer does not believe that any such errors occurred in this case, for three reasons. In the first place, computations were done with all the usual checks, and were then repeated independently; identical results were obtained the second time. This does not conclusively eliminate the possibility of computational errors, but makes their occurrence rather less likely.

In the second place, the argument presented by Rokeach and Hanley in favor of their view is a very indirect one, as the published article does not contain enough detail to make accurate computation possible. As they themselves admit, in discussing the "0" responses, "there is only indirect evidence bearing on this point. . . . "[1] It is, in fact, impossible to argue back from the published figures in the way that Rokeach and Hanley are doing, and no rigorous development of their criticism is indeed attempted. When they say of their "recomputations" that "in only two out of eight comparisons are the means identical," it should be clearly understood that this is quite irrelevant as their recomputations leave out part of the data. The fact that the means in two cases are identical is purely fortuitous; there is no reason why any of the means should be identical.

In the third place, Rokeach and Hanley have been very partial in their selection of evidence. They say that "in view of the foregoing analysis, Eysenck's continued contention that communists are more tough-minded than conservatives, liberals, and socialists, is not supported by his published data." Yet in the *Psychology of Politics* (**11**, p. 141), there is given a detailed diagram of the scores made by communists and fascists, as compared with a group of matched subjects of conservative, liberal, and socialist attitudes; this diagram bears out completely the conclusion criticized by Rokeach and Hanley. The figures on which it is based, contained in a doctoral dissertation by Coulter (**4**), were available to at least one of the two critics, and the published diagram gives sufficient detail to show that this independent research substantiates the contention that communists are more tough-minded than people supporting other political parties (with the exception of the fascists). The failure to mention this corroborative evidence is difficult to explain.

Equally important in this connection is another research, completed only recently, and as yet unpublished. This study by Nigniewitzky could not have been known to Rokeach and Hanley, but the results are very relevant to the question of whether the original data can be duplicated in repeated and independent studies. Basing his study on a properly stratified sample of the French population, and using a slightly modified and improved form of the T scale, Nigniewitzky found that communists had a mean score of 10.3; fascists had a mean score of 10.2; communist fellow-travellers had a mean score of 10.2. The mean score of supporters

[1] Even this "indirect evidence" of theirs is based on curious reasoning and factual inaccuracies. Thus Rokeach and Hanley say: "He [Eysenck] *always* scores "0" responses as tough-minded. . . ." This is quite untrue. Several different scoring schemes have been tried out at various times, such as the one mentioned in the 1947 paper (**6**, p. 65). The work of Melvin (**15**) has contributed greatly to a final decision on the best method of dealing with the problem of the "0" response. Any recomputation based on false assumptions of this kind must be regarded as irrelevant.

of all the other main French parties was 17.6! These figures are even more impressive than those found in England; they strongly support our view regarding the position of communists in the two-dimensional factor space.

One further result from Nigneiwitzky's study may be of interest. He found that in an analysis of variance carried out over the main political parties in France, the score on the T dimension gave an even better differentiation than did the R score (in Anglo-Saxon countries the opposite is usually found). Other scales, such as the F scale, which bears considerable similarity to the T scale,[2] and correlates reasonably highly with it in most studies, were very much inferior to both the T and the R scales. These facts, added to those reported in the *Psychology of Politics* may serve as an adequate comment on Rokeach and Hanley's contention that "tough-mindedness-tender-mindedness,' as conceived and measured by Eysenck, has no basis in fact. It is based on miscalculation and a disregard for a significant portion of his data. It conceals rather than reveals the attitudinal differences existing among political groups."

All in all, then, our answer to Rokeach and Hanley is that proper care was observed in the calculation of the data; that their criticism is not based on rigorous calculation, but on argument and surmise; and that two independent repetitions of the study, one of which was known to Rokeach and Hanley, give results even more striking in their support of our hypothesis than did the original study investigated by Rokeach and Hanley.

Allied to the criticism regarding the alleged computational errors is Rokeach and Hanley's discussion of the detailed results of the 1951 paper. They take the writer to task because "in coming to his conclusions (he) makes no reference whatever to these data on the individual items." This is the first time the writer has been criticized for obeying Rule 1.22, Subsection *d*, of the APA Publication Manual (**2**), which reads: "Data should be presented no more than once. Although it is appropriate to refer to tabular data in the text of an article, care should be taken not to repeat data unnecessarily in the section on results, in the discussion, and in the summary." The tabular presentation was sufficiently detailed for Rokeach and Hanley to draw conclusions from it at considerable length; no editor would have permitted the writer a discussion of similar length in addition to the tabulation. However, the main point of their discussion indicates that Rokeach and Hanley fail to understand the chief characteristic of dimensional analysis. Communists as a group have loadings on *two* orthogonal factors; consequently their responses to individual items are determined not only by their tough-mindedness, but also by their radicalism. Items relating to anti-Semitism, war attitudes, the death penalty, and so forth should be answered in the affirmative because of their loading on tough-mindedness.

[2] Historically the T scale was published several years before the F scale. The T dimension was isolated in 1944 (**5**), and the scale published in 1947 (**6**). The F scale was published in 1950 (**1**), without mention of the T scale in spite of the obvious similarities. Neither was Ferguson's (**13**) contribution mentioned, which also is very relevant to the concepts underlying the F scale. Rokeach and Hanley take the author to task because he "did not mention Ferguson's 1941 paper in his original publications on the R and T factors." They omit to add that in an even earlier paper, not quoted by them at all, the writer (**5**) had thoroughly and in detail discussed the contribution not only of Ferguson (**12**), but also of Carlson (**13**), Thurstone (**17**), and many others.

but in the negative because of their loading with radicalism; the outcome of the ensuing conflict will depend on the respective loadings, as well as on the exact position of each person in the communist group on the two factors. The T score combines in equal proportions radical and conservative items and thus gets rid of the complication introduced by the R factor; in just the same way the R score, combining in equal porportions tough-minded and tender-minded items, gets rid of the complications introduced by the T factor. This point appeared too obvious and indeed elementary to discuss at length in the paper; the reader interested in the detailed construction of the scales, and the problems encountered, may be referred to a separate publication by Melvin (**15**).

We may now turn to the second major criticism presented. In discussing the similarity between his dimensional scheme and that presented by Ferguson (**12**, **13**), the writer (**11**, p. 147) has commented that a rotation of 45° would turn the one pair of reference axes (T and R) into the other (humanitarianism and religionism). There is an obvious semantic convenience in employing widely used and accepted terms, such as radicalism-conservatism, particularly when there is evidence that the scale for measuring such a factor coincides with the actual major political party groupings (**6**). Furthermore, it seems more reasonable to refer to communists as "tough-minded radicals," or to fascists as "tough-minded conservatives," than to refer to conservatives as "religious antihumanitarians," or to socialists as "nonreligious humanitarians," as we would have to do if we accepted the Ferguson scheme. Indeed, this rechristening seems to lead to a *reductio ad absurdum* when we find Rokeach and Hanley arguing that "communists score the highest on the humanitarian items." To find communists considered as the most "humanitarian" group of all is certainly a little startling!

However, this argument regarding the superiority of the R and T dimensions on the basis of semantic convenience was only used by the author in a very subsidiary way. As pointed out in *The Psychology of Politics*, in a passage quoted by Rokeach and Hanley, "more convincing would be experimental evidence showing that Tough-mindedness had correlates in other fields, such as, for instance, in the field of personality, which neither Religionism nor Humanitarianism possessed. A proof of this type will be attempted in a later chapter . . . " (**11**, p. **147**). Rokeach and Hanley comment: "Such evidence would indeed be instructive. We made a careful search of the remainder of *The Psychology of Politics* for this promised experimental test. Our search was in vain. The issue is never again raised in the book." The writer finds this comment difficult to understand. A whole chapter, entitled Ideology and Temperament, is given over to a discussion of the experimental evidence relating to this problem, and several different approaches are reported, all of which support the hypothesis that tough-mindedness and extraversion are related to each other, as required by our hypothesis. The reader's attention is drawn particularly to Figure 30, on p. 178 of *The Psychology of Politics*, which reports the results obtained by George (**14**) in a direct attack on this problem. It will be seen there that his measure of extraversion is situated almost exactly on the tough-minded factor axes. Anyone

familiar with dimensional analysis will be able to see for himself the result of rotating the axes through an angle of 45°, thus bringing them in line with the Ferguson system. This would considerably reduce the correlation of extraversion from its present reasonably high size, and would leave us with two rather low and unimportant correlations with religionism (negative) and with humanitarianism (negative). Furthermore, the relation between extraversion and tough-mindedness observed in this study was predicted in terms of theoretical considerations; no such prediction was made to our knowledge with respect to Ferguson's two factors. Rokeach and Hanley's failure to see the relevance of this whole chapter, and of this study in particular, to the point in question is difficult to understand.

They also fail to take into account what to the writer is the most important chapter in the whole book, viz., the concluding chapter entitled "A Theory of Political Action." Here an attempt has been made to deduce the actual structure of attitudes found, as well as the relationship of the T factor to extraversion-introversion, from general learning theory; it was also deduced that there should be no consistent relationship between the R factor and the main personality variables. None of the relations pointed out in this chapter, and none of the deductions made, would be applicable to the Ferguson factors. Rokeach and Hanley do not mention this argument, although to the writer it appears the most cogent one in coming to a decision between the two rival schemes. This failure to come to grips with the writer's theory as a whole appears to him the outstanding weakness in the critique to which this is the reply. The authors have quite arbitrarily picked out certain isolated points, have disregarded the great mass of evidence supporting each separate conclusion, as well as the interconnections between the different parts of the research under review, and have come to conclusions which are not in fact borne out by a careful perusal of the evidence. The reader will be able to form his own opinion after comparing the facts as outlined in *The Psychology of Politics* with Rokeach and Hanley's critique.

REFERENCES

1. Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. A.P.A. Publication manual of the American Psychological Association. *Psychol. Bull. Suppl.*, 1952, **49**, 1–61.
3. Carlson, H. B. Attitudes of undergraduate students. *J. soc. Psychol.*, 1934, **5**, 202–212.
4. Coulter, T. An experimental and statistical study of the relationship of prejudice and certain personality variables. Unpublished doctor's dissertation, Univer. of London, 1953.
5. Eysenck, H. J. General social attitudes. *J. soc. Psychol.*, 1944, **19**, 207–227.
6. Eysenck, H. J. Primary social attitudes: 1. The organization and measurement of social attitudes. *Int. J. Opin. Attitude Res.*, 1947, **1**, 49–84.
7. Eysenck, H. J. *Dimensions of personality.* London: Kegan Paul, 1947.
8. Eysenck, H. J. Primary social attitudes as related to social class and political party. *Brit. J. Sociol.*, 1951, **2**, 198–209.
9. Eysenck, H. J. *The scientific study of personality.* London: Routledge & Kegan Paul, 1952.
10. Eysenck, H. J. *The structure of human personality.* London: Methuen, 1953.
11. Eysenck, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.

12. FERGUSON, L. W. Primary social attitudes. *J. Psychol.*, 1939, **8**, 217–223.
13. FERGUSON, L. W. The stability of the primary social attitudes. 1. Religionism and humanitarianism. *J. Psychol.*, 1941, **12**, 283–288.
14. GEORGE, E. I. An experimental study of the relations between personal values, social attitudes, and personality traits. Unpublished doctor's dissertation, Univer. of London, 1954.
15. MELVIN, D. An experimental and statistical study of two primary social attitudes. Unpublished doctor's dissertation, Univer. of London, 1955.
16. ROKEACH, M., & HANLEY, C. Eysenck's tender-mindedness dimension: a critique. *Psychol. Bull.*, 1956, **53**, 169–176.
17. THURSTONE, L. L. The vectors of mind. *Psychol. Rev.*, 1934, **41**, 1–32.

Received April 11, 1955.

Psychological Bulletin
Vol. 53, No. 2, 1956

CARE AND CARELESSNESS IN PSYCHOLOGY

CHARLES HANLEY AND MILTON ROKEACH
Michigan State University

Our joint reaction to Eysenck's reply (**6**) is that it evades or beclouds each of the specific issues we raised concerning his work on "tender-mindedness." To our mind, the very title of his reply is misleading. It is not *The Psychology of Politics* (**5**) we have criticized. The reader whose attention is directed to this book will not be able to see for himself the serious computational errors, the omission of contradictory evidence, and the erroneous conclusions we have pointed to in our critique.[1] Eysenck republishes in this book a good deal, but not all, of the data he presented in earlier journal articles. Most conspicuous is the omission of his data on the responses of communist and working-class samples to the individual items of his scale. He republishes such data only for middle-class conservatives, liberals, and socialists. It is precisely the omitted data which show that the communists are more "tender-minded" than conservatives, liberals, and socialists on 8 of the 14 T items, and it is these same data that enabled us to recompute mean T scores for his various groups and thereby discover that the means he reports are incorrect. It is for this reason that we felt it necessary to scrutinize Eysenck's earlier and, by far, fuller reports (**3**, **4**), rather than the data Ensenck has chosen to present in his book.

Now let us examine Eysenck's reply to each of our criticisms.

Concerning responses to individual T items. Eysenck states that by "dimensional analysis" it is possible to reconcile the fact that the communists, whom he describes as being tough-minded, turn out to be the most tender-minded of any group on 8 of the 14 T items. His reasoning runs somewhat as follows. Each of the 14 items not only measures individual differences along the tender-mindedness axis but also along the radicalism axis. Suppose that agreement with a specific item is scored both as tender-minded and as radical, and that the subject is a communist. Eysenck attributes such agreement to radicalism and not to tender-mindedness. He does not, however, consistently apply this line of thought to the other items. Thus, suppose that agreement with an item is scored both as tough-minded and radical, and the subject is a communist. If Eysenck were to follow his rule, such agreement should be attributed to radicalism and not to tough-mindedness. This he does not do. Instead, he attributes such agreement to the operation of tough-mindedness. Suppose, again, that agreement is scored both as tender-minded and conservative, and that the subject is a conservative. If Eysenck were consistent, he should interpret such agreement as being due to conservatism, and not to tender-mindedness. This time, however, such agreement is attributed to *both* conservatism and tender-mindedness.

We cannot escape the impression that following Eysenck's line of argument permits one to shift one's ex-

[1] A reading of this book reveals shortcomings and errors other than those we have considered. Many of these will be dealt with in a separate critique by Professor Richard Christie in a forthcoming issue of this journal (**1**).

planation of results from one axis to the other according to which axis one wishes to grind.

Eysenck asserts also that the reason he never referred to the contradictions in the item response frequencies is that he was prevented from doing so because he was "obeying Rule 1.22, Subsection *d*, of the APA Publication Manual," which is to the effect that data should not be repeated unnecessarily.

We find ourselves hesitant to take this explanation seriously. We will take space here only to point out that the APA Publication Manual Eysenck refers to was published in 1952. The article containing the contradictory findings appeared in 1951—in the *British Journal of Sociology!*

Concerning Eysenck's incorrect means. When we recomputed means for Eysenck's groups from the data he presents on response frequencies to items, we found serious discrepancies between his means and ours. We suggested that his means were in error to such an extent that his conclusions regarding differences in tough-mindedness among communists, socialists, liberals and conservatives had no basis in his data. Eysenck replies "computations were done with all the usual checks, and were then repeated independently." He denies the correctness of our recomputed means because, he says, it is "impossible" to do so using the response frequencies. These recomputations, he continues, are based on "false assumptions" and are "irrelevant," because we do not know how the "0" responses were scored. We are referred to the work of Melvin who "has contributed greatly to a final decision on the best method of dealing with the problem of the '0' response."

Several comments are in order here.

1. The most convincing way to demonstrate the fallaciousness of a critic's recomputations would be to make available the raw data and detailed instructions on how they are to be scored. Eysenck has made no effort to do this.

2. We do not believe that Eysenck's conclusions are worthy of serious attention if different ways of scoring "0" responses can make such whopping differences in findings.

3. Eysenck himself describes on page 276 of *The Psychology of Politics* exactly how Melvin scores the T scale. As far as one can tell from Eysenck's description (**5**, p. 65), his original scoring method is identical with Melvin's.

4. However the "0" responses may have been handled, the crucial data in Eysenck's table of response frequencies to individual T items are already classified into proportions of "Yes" answers, and the accompanying text describes the scoring system in terms of "Yes" answers to items (**4**, pp. 201, 203). Thus, if his means fail to correspond with the means which anyone may calculate from the data presented in this table, Eysenck's means must be incorrect.

5. We have noted previously the carelessness which Eysenck has shown in presenting factual material. Let us note two further examples from his reply.

First, Eysenck gives in his reply the year 1955 as the date of Melvin's Ph.D. thesis. In *The Psychology of Politics* reference is also made to this thesis. On page 276 the date given is 1954; and on page 301 the date is 1953! (We note too that Eysenck refers to this *unpublished* dissertation as "a separate publication by Melvin.")

Second, Eysenck writes in his reply: "Rokeach and Hanley take the author to task because he 'did not mention Ferguson's 1941 paper in his original publications on the R and T

factors.' They omit to add that in an even earlier paper, not quoted by them at all, the writer (**5**) had thoroughly and in detail discussed the contribution not only of Ferguson (**12**), but also of Carlson (**13**), Thurstone (**17**), and many others."

A careful consideration of this statement shows that "Ferguson's 1941 paper" and the "Ferguson (**12**)" reference ***are not the same paper!*** The "Ferguson (**12**)" cited by Eysenck turns out to be another study by Ferguson published in 1939.[2] We suggest that if one now re-reads the above quotation it will have an entirely different meaning!

These two examples demonstrate marked carelessness by Eysenck in assembling and presenting factual material. Application of some of the "usual checks," which Eysenck states he employed in calculating his means, would also have prevented these errors from occuring.

Concerning the independent repetitions of Eysenck's research. Eysenck states that two independent repetitions of his original study by Coulter and Nigniewitzky "give results even more striking in their support of our hypothesis."

The internal consistency and computational correctness of a particular investigator's data cannot be established by referring to two or even more than two independent studies. On the basis of our evaluation of Eysenck's published research, we have come to the conclusion that his data do not support his hypothesis. Whether the two unpublished studies of his students confirm or deny Eysenck's hypothesis is an entirely separate issue. These studies deserve evaluation strictly on their own merit. We look forward to the publication of these reports.[3]

Concerning the alleged superiority of Eysenck's factors over Ferguson's factors. We pointed out that one may more easily understand Eysenck's results by referring to Ferguson's "Religionism" and "Humanitarianism" factors rather than to Eysenck's R and T. We called attention to the fact that Eysenck promises in *The Psychology of Politics* to produce "experimental evidence showing that Tough-mindedness had correlates in other fields, such as, for instance, in the field of personality which neither Religionism nor Humanitarianism possess."

We then pointed out that Eysenck never again discusses this issue. The matter is simply dropped! Eysenck replies that he finds our "comment difficult to understand. A whole chapter, entitled 'Ideology and Temperament,' is given over to a discussion of the experimental evidence relating to this problem, and several different approaches are reported, all of which support the hypothesis that tough-mindedness and extraversion are related to each other, as required by our hypothesis."

The only way possible to demonstrate the proof he promises of the superiority of his factors over Ferguson's is by a *pitting* procedure. It is not sufficient to present *only* the correlations between T and selected personality variables. It is necessary to produce, in addition, the comparable correlations between Ferguson's factors and the same personality variables. Only then can one make a choice between the two alternative explanations. This is the test Eysenck promises but fails to make.

There is no reference whatever in

[2] Involved here is more than a careless substitution of one reference by another. Ferguson's 1941 paper describes a Religionism factor that Eysenck in 1944 (**2**) complained was missing from Ferguson's 1939 study. Eysenck's 1944 paper makes no mention of Ferguson's 1941 paper.

[3] A detailed analysis of the deficiencies in Coulter's study, as reported by Eysenck, will be found in Christie's paper (**1**).

his chapter on Ideology and Temperament (or in his Figure 30, for that matter)[4] to the Ferguson factors or their correlates. Indeed, nowhere does he suggest that he has computed scores on these dimensions and correlated them with the relevant personality variables. We, therefore, must reaffirm our earlier statement: "Such evidence would indeed be instructive. We made a careful search of the remainder of *The Psychology of Politics* for this promised experimental test. Our search was in vain. The issue is never again raised in the book."

A final suggestion. If one is interested, for the sheer fun of it, in confirming Eysenck's conclusion that fascists and communists are similar to each other and different from democratic subjects, we offer the following recipe. Construct a 20-item scale. Include 10 items referring to acceptance of communist ideology (e.g., "Communism is the most desirable form of government," etc.). Let the remaining 10 items refer to acceptance of fascist ideology (e.g., "Fascism is the most desirable form of government," etc.). Give the questionnaire to communists, socialists, liberals, conservatives, and fascists. Factor items. Emerge with two factors. Call one, "radicalism," the other, "tough-mindedness." Score agreement with communist items and disagreement with fascist items as "radical." Discover that communists are the most radical, fascists the most conservative, and democratic groups in between. Score agreement with both communist and fascist items as "tough-minded." Discover that communists and fascists are both "tough-minded," because they agree with 10 of the 20 items. Find democratic groups to be "tender-minded," because they agree with none of the items.

In our opinion, Eysenck is caught in precisely this sort of trap. We hope, as a result of this exchange, that others will be able to steer clear of it.

[4] Eysenck (**6**) asserts that anyone can see from his Figure 30 (**5**, p. 178) that rotating 45° to the Ferguson axes reduces "the correlation of extraversion from its present reasonably high size, and would leave us with two rather low and unimportant correlations with religionism (negative) and humanitarianism (negative)." Following this *ad hoc* suggestion, we made the necessary rotation on Figure 30, measured the loading of extraversion on the new Religionism axis, and found that it is .88 of the magnitude of extraversion's loading on the T axis. The text (**5**, p. 179) states that the correlation between extraversion and tender-mindedness is .41. Thus, the estimated correlation between extraversion and Religionism is 88 per cent of .41 (negative), or −.36. Readers may wonder why Eysenck should consider a correlation of .41 as "reasonably high," but a correlation of .36 as "rather low and unimportant."

REFERENCES

1. Christie, R. Eysenck's treatment of the personality of communists. *Psychol. Bull.* in press.
2. Eysenck, H. J. General social attitudes. *J. soc. Psychol.*, 1944, **19**, 207–227.
3. Eysenck, H. J. Primary social attitudes: I. The organization and measurement of social attitudes. *Int. J. Opin. Attitude Res.*, 1947, **1**, 49–84.
4. Eysenck, H. J. Primary social attitudes as related to social class and political party. *Brit. J. Sociol.*, 1951, **2**, 198–209.
5. Eysenck, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.
6. Eysenck, H. J. The psychology of politics: a reply. *Psychol. Bull.*, 1956, **53**, 177–182.
7. Rokeach, M., & Hanley, C. Eysenck's tender-mindedness dimension: a critique. *Psychol. Bull.*, 1956, **53**, 169–176.

Received November 17, 1955.

Out-Of-Print Journals

☆ ☆ ☆

The American Psychological Association often receives requests for out-of-print items. The cost of reprinting these items would be justified if the demand were great enough. It would, therefore, be helpful in making the decision to reprint a particular item if those interested would send in a *tentative* order. If you would like to have a specific journal or *Psychological Monograph* that is now out of print, please complete and send in the following form.

Publications Office
AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N.W.
Washington 6, D.C.

If ....
(name of journal, volume, and number)

is reprinted, I would order a copy.

Name ....

Address ....

City ....Zone ......State ....

APA Members and Journal Subscribers—Are you going to move?

If you move—

your journals will not follow you from your old address to your new one

When you move—

notify the APA Subscription Office

Formerly, journals that could not be delivered because subscribers had not notified the APA of a new address were reclaimed by the APA, and the journal was remailed to the subscriber at his new address. This was always expensive. Recent changes in the postal laws and regulations have made the expense prohibitive. Undeliverable copies are now destroyed by the Post Office. Subscribers who do not receive a journal because of an address change are charged the regular single issue price for a replacement copy.

So—when you move—

Notify the postmaster at your old address and guarantee that you will pay the forwarding postage.

Notify the APA Subscription Office as early as possible—by at least the fifteenth of the month preceding the month when the change should take effect.

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N.W.
Washington 6, D.C.

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN